The Report Committee for Kyle Fred Scheele

Certifies that this is the approved version of the following report:

Wind Forecast Verification:

A Study in the Accuracy of Wind Forecasts Made by The Weather Channel and AccuWeather

APPROVED BY

SUPERVISING COMMITTEE:

Supervisor: ___________________________________

Eric Bickel

____

John Hasenbein

Wind Forecast Verification:

A Study in the Accuracy of Wind Forecasts Made by The Weather Channel and AccuWeather

by

Kyle Fred Scheele, B.S.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Engineering

The University of Texas at Austin

August 2011

Acknowledgements

I gratefully acknowledge and thank Eric Floehr and Intellovations, LLC for providing the dataset that was used within this report. I also heartily thank my supervisor, Eric Bickel, whose supervision and support from the preliminary to the concluding level enabled me to develop an understanding of the subject.

Abstract

Wind Forecast Verification:

A Study in the Accuracy of Wind Forecasts Made by The Weather Channel and AccuWeather

Kyle Fred Scheele, M.S.E.

The University of Texas at Austin, 2011

Supervisor: Eric Bickel

The Weather Channel (TWC) and AccuWeather (AWX) are leading providers of weather information to the general public. The purpose of this Master's Report is to examine the wind speed forecasts made by these two providers and determine their reliability and accuracy. The data used within this report was collected over a 12-month period at 51 locations across the state of Texas. The locations were grouped according to wind power class, which ranged from Class 1 to Class 4. The length of the forecast period was 9 days for TWC and 14 days for AWX.

It was found that the values forecasted by TWC were generally not well calibrated, but were never far from being perfectly calibrated and always demonstrated positive skill. The sharpness of TWC's forecasts decreased consistently with lead time, allowing them to maintain a skill score greater than the climatological average throughout the forecast period. TWC tended to over-forecast wind speed in short term forecasts, especially within the lower wind

power class regions. AWX forecasts were found to have positive skill the first 6 days of the forecasting period before becoming near zero or negative. AWX's forecasts maintained a fairly high sharpness throughout the forecast period, which helped contribute to increasingly un-calibrated forecast values and negative skill in longer term forecasts. The findings within this report should help provide a better understanding of the wind forecasts made by TWC and AWX, and determine the strengths and weaknesses of both companies.

Table of Contents

List of Tables

List of figures

Chapter 1: Introduction

The field of forecast verification got its modern start when Finley (1884) analyzed data on whether or not a tornado would occur. While this marked the beginning of the modern era of forecast verification, there would not be a substantial interest in the field for several decades. It wasn't until the advent of numerical weather forecasting, and the accompanying expansion of weather forecast products in the 1950's, that there was a significant expansion in the research effort to evaluate the validity of the forecasts being made. In the years and decades that followed, many important findings would be made within the field. One of the most important of these findings was made by Murphy and Winkler (1987), when they established a general framework for forecast verification based on the joint distribution of events that still influences much of the research being done today.

In the years following 1950, little attention was paid to wind forecast verification. This would change considerably in the 1970's with the growth of wind power as a more viable energy option, coupled with the overall expansion of wind measurement resources. In recent years there have been several categories of research being done with regards to wind forecasts. One such category is within the framework of an overall forecasting model. In addition to wind forecasts, these models generally include forecasts for characteristics such as precipitation, temperature, dew point, and relative humidity. Generally, the models of this type that have been evaluated are either from government agencies or research institutions (White et al. 1998).

Another area of interest to researchers in recent years has been to experiment and expand upon the current models in use. Researchers like Feddersen and Sattler (2005) have attempted to find more accurate models through the use of advanced techniques and increased computing power, while others have been attempting to verify off-shore wind forecasts, such as over the Mediterranean Sea (Accadia et al. 2006). The last area we will discuss deals specifically with the value of accurate forecasts to the wind energy field. Much of the work done in this field focuses on cost-benefit analysis and determining what a company should be willing to pay for an accurate wind forecast, rather than on the actual verification of wind forecasts. However, research in this field illustrates the value of improving the accuracy and reliability of wind forecasts (Milligan et al. 1995). While there has been significant research done in the areas discussed above, limited research has been done looking into the accuracy and reliability of the forecasts of commercial providers, and it is this area that this report will focus on.

Company Background

The Weather Channel (TWC) was launched on May 2, 1982, and since then has become a leading provider of weather information to the American public, and has begun expansion into international markets. TWC reaches the public through its cable television network, terrestrial and satellite radio stations, newspapers, and their interactive website (see http://www.weather.com). The cable network is received by more than 99 million households and can be seen in more than 97% of all cable TV homes nationwide. The interactive website, referred to as TWCi, provides forecasts for more than 100,000 locations worldwide and reaches more than 41 million unique users online each month and is the most popular source of online weather, news and information according to the Nielsen//NetRatings (additional information is available online at http://press.weather.com/company.asp).

Founded in 1962, AccuWeather (AWX) is an American company that provides for profit weather forecasting services worldwide. AWX provides local forecast information for the entire

United States, as well as many regions worldwide. They provide their products and services to more than 175,000 paying customers in media, business, government, and institutions. Their forecasts can be found in a wide variety of media, including newspapers, radio, television, and online at their free online weather provider AccuWeather.com. Their television network, AccuWeather Network, is now available in over 50 broadcast markets plus 48 cities through AT&T U-verse as an interactive channel (additional information is available online at http://www.accuweather.com/company.asp).

The general public uses the wind forecasts of outlets such as TWC and AWX to help make decisions regarding their day-to-day outdoor activities. In addition to the planning of the mundane, wind forecasts are also of great importance in the operation and maintenance of wind farms. The importance of wind forecasts to a field as large as the wind power industry, where the difference between an accurate forecast and an inaccurate one can cost them millions of dollars over the course of a year, raises several important questions. How accurate are these forecasts? Are these forecasts free from bias? How far into the future can these forecasts be seen to contain useful information? These are the questions we will attempt to answer within this report.

In this paper, we analyze the reliability of wind speed forecasts provided by TWC and AWX over a 12-month period (3 May 2010 – 30 April 2011), at 51 locations across Texas. Specifically, we compare *n*-day-ahead wind forecasts, where *n* ranges from 1 to 9 days for TWC, and *n* ranges from 1 to 14 for AWX, with actual wind speed observations. Additionally, the 51 locations will be sorted according to wind power class, for which the regions in question range from Class 1 to Class 4. The specifics of the wind power classifications will be discussed in the following section.

This report is divided into six chapters. In the next chapter, we describe the wind power classification system and examine the different power type regions within Texas. In chapter 3 we describe our verification approach. In chapter 4 we summarize the data collected for this analysis. In chapter 5 we present the verification and reliability results of the analysis and discuss their implications. Finally, in chapter 6 we present our conclusions.

Chapter 2: Wind Power Classification

In 1973, the United States faced an energy crisis created by the Arab oil embargo. This crisis led the United States government to take a serious interest in wind energy for the first time, and it was during this period that the assessment of national wind resources became a top priority. Early research in wind characteristics included the development of techniques for estimating the magnitude and distribution of wind over a selected area. In 1979, the Pacific Northwest Laboratory (PNL) used these assessment techniques in preparing twelve regional wind energy atlases covering the United States and its territories.

The initial assessment contained some areas with limited or no data that had to be estimated through the examination of nearby sites. In the years that followed, hundreds of new sites would go up across the nation specifically designed for wind energy assessment. In 1983, the Department of Energy administered a program run by the PNL to identify and assimilate the new data from these sites. This new data would allow for the verification and correction of early estimates of wind power in areas that did not have sufficient data in the past. With this expanded data set, the PNL was able to update and improve its regional wind energy atlases. The wind power class data used in this report comes from these updated maps.

Each wind power class represents the range of wind power densities and wind speeds likely to be encountered at exposed sites within an area designated as having that wind power class. The ranges of wind speed and wind power density for each wind power class are shown in Table 1. From the table, it can be seen that the average wind speed at 50m (the average height of a wind monitoring tower) within a wind power class 1 region would be expected to range from 0 to 12.5 mph, the average wind speed for a wind power class 2 region would range from 12.5 to 14.3 mph, and so forth. The wind power density is calculated using the air density and *n* different wind speed observations. We will only be examining the wind speed within this report, so further description and analysis of wind power density will not be given.

Table 1: Classes of wind power density at 10 m and 50 m

Wind Power Class	10 m (33 ft)		50 m (164 ft)	
	Wind Power Density (W/m^2)	Speed m/s (mph)	Wind Power Density (W/m2)	Speed m/s (mph)
1	0	0	0	0
	100	4.4 (9.8)	200	5.6 (12.5)
2				
	150	5.1 (11.5)	300	6.4 (14.3)
3				
	200	5.6 (12.5)	400	7.0 (15.7)
4				
	250	6.0 (13.4)	500	7.5 (16.8)
5				
	300	6.4 (14.3)	600	8.0 (17.9)
6				
	400	7.0 (15.7)	800	8.8 (19.7)
7				
	1000	9.4 (21.1)	2000	11.9 (26.6)

Based upon the criteria within Table 1, the PNL constructed regional maps for all of the 50 states. The regional map for Texas was split into two parts due to its size. These maps are the ones used for the classification of locations by wind power class within this report, and they are shown in Figure 1.

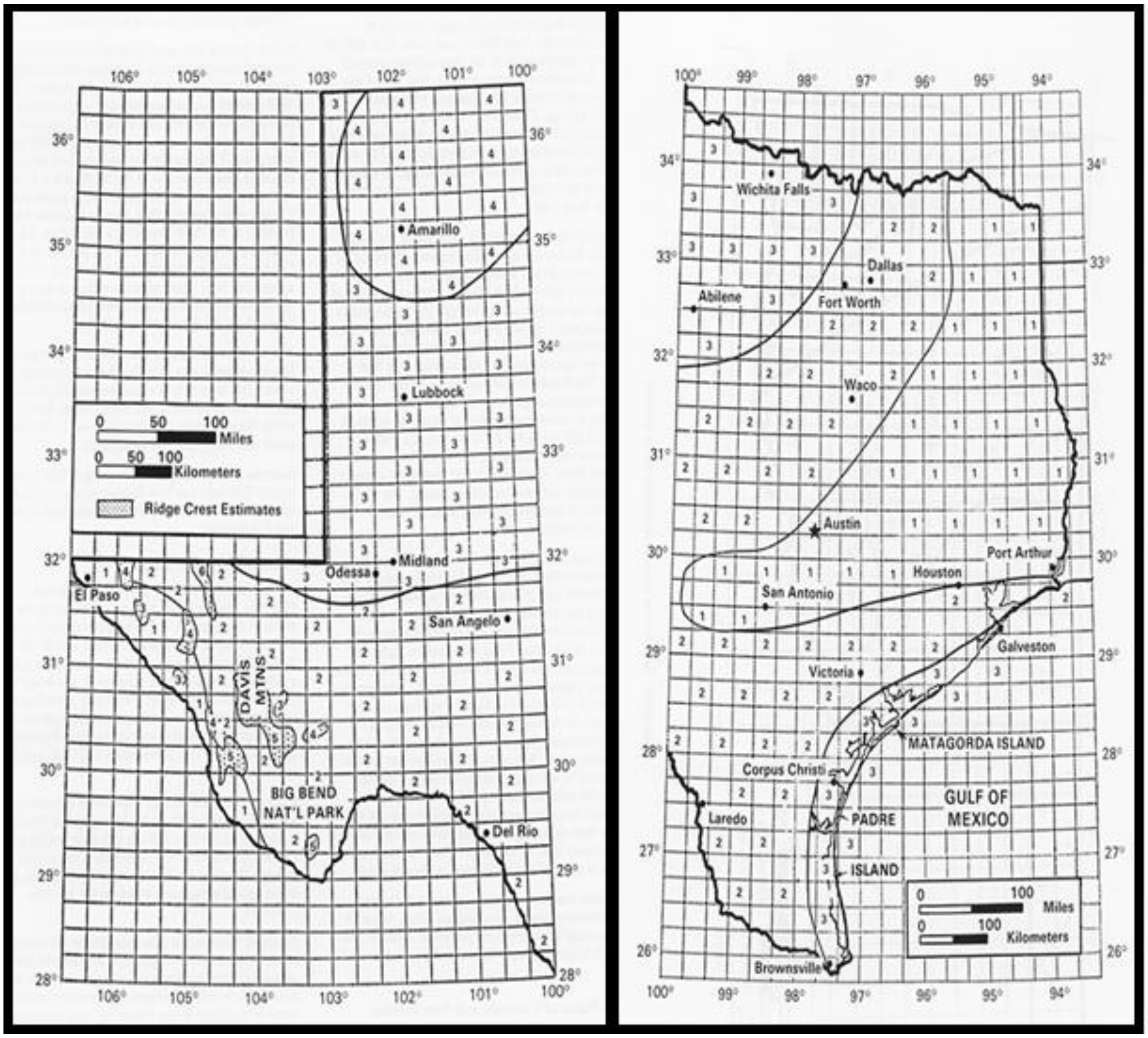


Figure 1: To the left is the PNL map of West Texas and to the right is the PNL map of East Texas.

Chapter 3: Verification of Continuous Value Forecasts

The literature dealing with the analysis and verification of forecasts is quite extensive. Many of the analysis general analytical techniques that will be used within this paper come from Jolliffe and Stephenson (2003). In addition, we make use of the distribution oriented framework proposed by Murphy and Winkler (1987), and the skill scoring system proposed by Murphy and Epstein (1989). The arrangement of this section is similar to Bickel and Kim (2008).

Distributional Measures

Let *F* be a finite set of possible wind speed forecasts $f_i \in [0, \infty), i = 1$ to m. Here *X* is the set of wind speed observations $x_i \in [0, \infty), i = 1$ to m. While there is no actual upper bound on either the potential values forecast or observed, a wind speed forecast or observation over 50 mph is rare. The probability mass function for the joint distribution of the forecasts and observations given a particular lead time is $p(f, x|l)$, and it gives the probability that the forecast f has a certain value at the same time that the observation x has a certain value. Therefore, a perfect forecasting system would be one that guaranteed that $p(f, x|l) = 0$ when $f \neq x$ for all l. The lead time l may take integer values ranging from 1 to 9 for TWC and from 1 to 14 for AWX.

The conditional distribution of the observations given the forecasts and the lead time is expressed as $p(x|f, l) = \frac{p(f,x|l)}{p(f|l)}$. This equation leads to the factorization of $p(f, x|l) = p(x|f, l)p(f|l)$, which is known as the calibration-refinement factorization. The term $p(x|f, l)$ is the conditional distribution of the wind speed observations given a particular forecast value and lead time, and the term $p(f|l)$ is the marginal distribution of forecasts for a given lead time. The forecasts and observations are independent if and only if $p(x|f, l) = p(x|l)$. A set of forecasts is said to be perfectly calibrated if $E[X|F=f]=f$ for all *f,* where E[ ] is the expectation operator. A set of forecasts is perfectly refined, or sharp, if $p(f|l)$ is equal to the marginal distribution of the observations, $p(x|l)$ for all *f, x,* and *l*. This would indicate that the spread of the forecasts is the same as the spread of the observations. Sharpness is important in differentiating between multiple well calibrated forecasting systems and vice versa. This relationship is illustrated through the following examples: Forecasting the climatological average will be well calibrated, but not sharp. Perfectly sharp forecasts are unlikely to be well calibrated.

The conditional distribution of the forecasts given the observations and the lead time is expressed as $p(f|x,l)=\frac{p(f,x|l)}{p(x|l)}$. This equation leads to the factorization of $p(f,x|l)=p(f|x,l)p(x|l)$, and is known as the likelihood-base rate factorization. The term, $p(x|l)$, represents the frequency distribution of wind speed observations. The term $p(f|x,l)$ is known as the likelihood function (or the discrimination), and it contains the information of what was forecast given a specific observed wind speed value. A forecasting system that forecasted the climatological average would have identical likelihood functions for each observed value. A perfect forecasting system would have a unique likelihood function for each observed value. This illustrates that the likelihood functions should vary across possible observations within a strong forecasting system. Lastly, the forecasts and observations are independent if and only if $p(f|x,l)=p(f|l)$.

Summary Measures

In addition to the distributional measures discussed in the previous section, we will also make use of several summary measures of forecast performance. The first of these is mean forecast value given a particular lead time, which is given by

$$\bar{f_l} = \sum_F \sum_X f p(f\,, x|l) = E_{F,X|l}[f]. \tag{1}$$

Similarly, the mean value for wind observations given a particular lead time is

$$\bar{x}_l = \sum_F \sum_X x p(f\,, x|l) = E_{F,X|l}[x]. \tag{2}$$

Mean error (ME), or bias, is a measure of the unconditional bias of a forecast and is given by

$$ME(f, x|l) = \bar{f_l} - \bar{x}_l. \tag{3}$$

Mean-square error (MSE) is one of the most commonly used forecast scores and is represented by

$$MSE(f, x|l) = \sum_F \sum_X (f - x)^2 = E_{F,X|l}\,[(f - x)^2]. \tag{4}$$

The mean-square error skill score (MSESS) compares the MSE of a forecast with the MSE of a climatological forecast to determine the forecast's effectiveness. The MSESS is

$$MSESS(f, x|l) = 1 - \frac{MSE(f,x|l)}{MSE(\bar{x},x|l)}, \tag{5}$$

Note that the denominator can also be expressed as

$$MSE(\bar{x}, x|l) = E_{F,X|l}[\,(\bar{x} - x)^2] = \sigma_x^2, \tag{6}$$

where σ_x^2 is the variance of the observations. Therefore, we can see that the MSESS serves as a way to measure the proportional reduction in variance that occurs as a result of the forecasting system.

Additionally, we will examine the correlation between the forecasts and observations, which is given by

$$\rho(f, x|l) = \frac{cov(f,x|l)}{(\sigma_x^2 \sigma_f^2)^{1/2}}, \quad (7)$$

where *cov* is the covariance of the forecasts and the observations and the denominator is the standard deviation of the forecasts times the standard deviation of the observations.

Lastly, we will assess the discrimination through the following measure:

$$DIS = var_X[E_F(F|X = x)],$$

(8)

where the right hand side represents the variance of the expected value of the forecasts, given a certain wind speed observation.

Chapter 4: Data gathering and data summary

The forecast data used for this project was collected daily from TWC and AWX websites by Eric Floehr and ForecastWatch. Forecasts were collected from both providers from 3 May 2010 to 30 April 2011. Forecast data was collected from a total of 51 locations across the state of Texas. Additionally, the corresponding observation data was collected from each location. These locations were provided in the data set along with the coordinates of their wind monitoring stations, allowing us to determine their wind power class.

The forecast elements that were collected daily were wind speed and wind direction for each lead time. The observation elements that were collected daily were four different measures of wind speed and wind direction. Upon inspection of the forecasts compared with the various measures, it was determined that the metric that is used within the forecasts of both providers was the 12-h average wind speed. This forecast covers the 12-h window between 0700-1900 local time, rather than covering a full 24-h day.

The collection of forecast data from TWC was successful nearly 100% of the time. The collection of forecast data from AWX was also largely successful, but it did have a higher failure rate than TWC collection process. We were unable to retrieve wind forecasts roughly 4% of the time from the AWX website. This data was excluded from the analysis. While wind direction data was collected by ForecastWatch, it was not included in our analysis.

Data Summary

Now that we have explained the data collection procedure, we can summarize the forecast and observation data that was obtained through Tables 2 and 3.

Table 2: Summary of forecast and observation data for TWC for all Classes.

Lead time (days)	No. of forecasts	Avg. wind speed observed (mph)	Avg. wind speed forecast (mph)	ME (Bias)
Class 1 TWC				
1	4956	10.686	12.188	1.502
2	4939	10.689	11.850	1.161
3	4925	10.696	10.042	-0.654
4	4907	10.689	9.842	-0.847
5	4896	10.667	9.644	-1.022
6	4888	10.685	9.438	-1.247
7	4870	10.660	9.256	-1.404
8	4855	10.649	9.208	-1.441
9	4839	10.636	9.176	-1.460
Class 2 TWC				
Lead time (days)	No. of forecasts	Avg. wind speed observed (mph)	Avg. wind speed forecast (mph)	ME (Bias)
1	6720	12.212	13.063	0.851
2	6702	12.196	12.753	0.557
3	6684	12.218	11.282	-0.936
4	6664	12.198	11.088	-1.110
5	6646	12.190	10.886	-1.304
6	6631	12.201	10.746	-1.455
7	6607	12.177	10.520	-1.657
8	6586	12.157	10.509	-1.648
9	6568	12.137	10.486	-1.651
Class 3 TWC				
Lead time (days)	No. of forecasts	Avg. wind speed observed (mph)	Avg. wind speed forecast (mph)	ME (Bias)
1	4097	14.008	14.700	0.692
2	4087	14.012	14.427	0.415
3	4078	14.015	13.145	-0.871
4	4066	14.005	12.881	-1.124
5	4053	14.008	12.770	-1.239
6	4039	13.989	12.610	-1.379
7	4027	13.974	12.423	-1.551
8	4014	13.939	12.409	-1.530
9	4005	13.926	12.462	-1.464
Class 4 TWC				
Lead time (days)	No. of forecasts	Avg. wind speed observed (mph)	Avg. wind speed forecast (mph)	ME (Bias)
1	1032	15.268	15.679	0.411
2	1029	15.289	15.547	0.258
3	1026	15.277	14.553	-0.724
4	1024	15.248	14.320	-0.929
5	1020	15.264	14.185	-1.078
6	1017	15.221	13.977	-1.244
7	1016	15.194	13.852	-1.342
8	1011	15.165	13.752	-1.413
9	1009	15.165	13.771	-1.394

Table 3a: Summary of forecast and observation data for AWX for Class 1 and 2 locations.

Class 1 AWX				
Lead time (days)	No. of forecasts	Avg. wind speed observed (mph)	Avg. wind speed forecast (mph)	ME (Bias)
1	4864	10.859	9.826	-1.034
2	4869	10.815	9.432	-1.384
3	4921	10.719	10.789	0.070
4	4907	10.694	11.105	0.411
5	4896	10.694	11.286	0.592
6	4891	10.709	11.367	0.659
7	4632	10.815	9.033	-1.782
8	4420	10.863	8.470	-2.393
9	4401	10.798	8.177	-2.622
10	4317	10.816	8.334	-2.482
11	4328	10.776	8.350	-2.426
12	4230	10.782	8.353	-2.428
13	4232	10.766	8.116	-2.650
14	3627	10.718	7.797	-2.921
Class 2 AWX				
Lead time (days)	No. of forecasts	Avg. wind speed observed (mph)	Avg. wind speed forecast (mph)	ME (Bias)
1	6714	12.229	11.141	-1.088
2	6697	12.205	10.846	-1.359
3	6669	12.216	12.390	0.174
4	6651	12.203	12.579	0.376
5	6646	12.200	12.782	0.583
6	6631	12.205	13.112	0.906
7	6467	12.249	10.353	-1.897
8	6281	12.274	9.699	-2.575
9	6272	12.256	9.428	-2.828
10	6181	12.229	9.707	-2.522
11	6115	12.245	9.715	-2.530
12	6033	12.275	9.590	-2.685
13	6022	12.181	9.342	-2.839
14	5072	12.202	9.196	-3.006

Table 3b: Summary of forecast and observation data for AWX for Class 3 and 4 locations.

Lead time (days)	No. of forecasts	Avg. wind speed observed (mph)	Avg. wind speed forecast (mph)	ME (Bias)
		Class 3 AWX		
1	4098	14.010	12.947	-1.063
2	4086	14.011	12.744	-1.267
3	4074	14.015	14.031	0.017
4	4065	14.004	14.056	0.052
5	4056	14.013	14.314	0.301
6	4043	13.999	14.598	0.598
7	3992	14.024	11.811	-2.213
8	3917	13.980	11.161	-2.819
9	3905	13.989	10.878	-3.111
10	3895	13.995	11.464	-2.531
11	3855	13.974	11.462	-2.512
12	3824	14.001	11.413	-2.588
13	3810	13.926	10.949	-2.977
14	3232	13.963	10.936	-3.026
		Class 4 AWX		
Lead time (days)	No. of forecasts	Avg. wind speed observed (mph)	Avg. wind speed forecast (mph)	ME (Bias)
1	1035	15.289	14.445	-0.844
2	1032	15.275	14.064	-1.212
3	1030	15.279	15.700	0.421
4	1026	15.258	15.170	-0.088
5	1023	15.259	15.800	0.541
6	1020	15.241	16.347	1.106
7	1015	15.246	12.096	-3.150
8	996	15.188	11.184	-4.003
9	993	15.176	11.085	-4.091
10	984	15.252	11.299	-3.953
11	982	15.160	11.114	-4.046
12	984	15.105	10.832	-4.273
13	961	15.084	10.106	-4.978
14	822	15.179	10.416	-4.764

A total of 149,534 usable forecast and observation pairs were collected from TWC during this period, and a total of 220,471 were taken from AWX. The number of forecasts collected from TWC for each lead time varied between 1,009 and 6,720 wind speed forecasts, depending on wind power class. For AWX the range was between 822 and 6,697. From the site location data we were able to determine the wind power class of each location. Out of the 51 locations, three were Class 4, twelve were Class 3, twenty were Class 2, and sixteen were Class 1. The average observed wind speed varied from over 10.5 mph in class 1 areas, to approximately 15.25 mph for class 4 areas. The class 2 and class 3 areas had observed wind speeds of approximately 12.2 mph and 14 mph respectively.

It can be seen that TWC wind speed forecasts attain a maximum forecast value at the 1-day lead time for all 4 Classes. Also, it can be seen that TWC forecasts achieved their minimum value at either 8 or 9 day lead time, depending upon the wind power class. The mean error for these forecasts is positive for 1 and 2 day lead times and then negative for all others for all wind power classes. Additionally, this summary also shows that the lower the wind power class, the greater the positive bias is for the first 2 days of the forecast. This means that TWC tends to over-predict the wind speed more for short term forecasts in regions with lower climatological average wind speeds. The full extent of this trend is illustrated through TWC's forecasts for the class 1 regions, where the bias is actually greater for the 1-day forecasts than for the 9-day forecasts. This style of forecasting suggests some degree of positive bias for the short term forecasts followed by some degree of negative bias for forecasts 3 days or more out.

For the AWX wind speed forecasts, the maximum value is attained at the 6 day lead time for all 4 Classes. The minimum value is achieved at either the 9, 13, or 14 day lead time, depending upon the wind power class. The mean error for AWX is seen to be positive for 3 to 6 day lead times (except for being slightly negative for 4 day lead time of class 4), and then negative otherwise for all wind power classes. This pattern suggests some degree of variation within the forecasting strategy of AWX within their 14-day forecasts.

Lastly, we will examine the distribution of forecasted values for both providers. In Figure 2, we can see the distribution of TWC forecasts for a given lead time, and in Figure 3, we can see the distributions of the AWX forecasts. For TWC, 11 mph is the most forecasted value for the 1-day lead time, and then this value decreases to 8 or 9 mph depending on lead time. The distribution of forecasted values smoothes out more and more for larger lead times, until by the end of the forecasting period the forecasts appear to be normally distributed around a mean of approximately 9 mph, which is slightly less than the average observed value from Table 2.

For AWX, 8 mph is the most forecasted value for the 1-day lead time, and then this value varies between 5 and 9 mph throughout the remainder of the forecast period. It can be seen that for the first few days of their forecasts, there is an indentation within the distribution between 12 and 13 mph, suggesting some kind of aversion to those values within their forecasting framework. Figure 3 also shows us that AWX does not forecast wind speed values greater than 0 but less than 4. These forecasts were rare for TWC, but were made occasionally. As was the case with TWC, the forecasts of AWX appear to smooth out as the lead time increases. Eventually the AWX forecast distribution appears to be similar to a lognormal distribution with a mean at just over 6 mph. Only the graphs for wind power class 1 for both providers are included in this section. The distribution graphs for all other wind power classes can be found in the appendix.

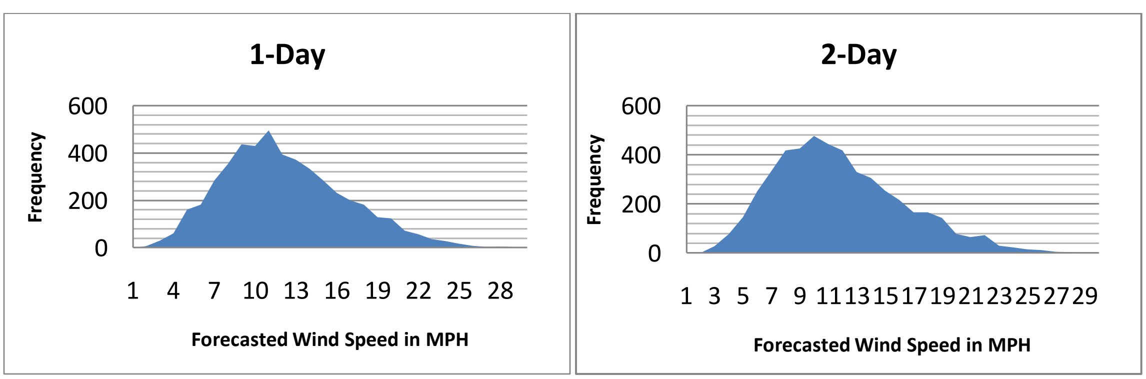


Figure 2a: TWC forecast distribution graphs for 1-2 day lead time for class 1 locations.

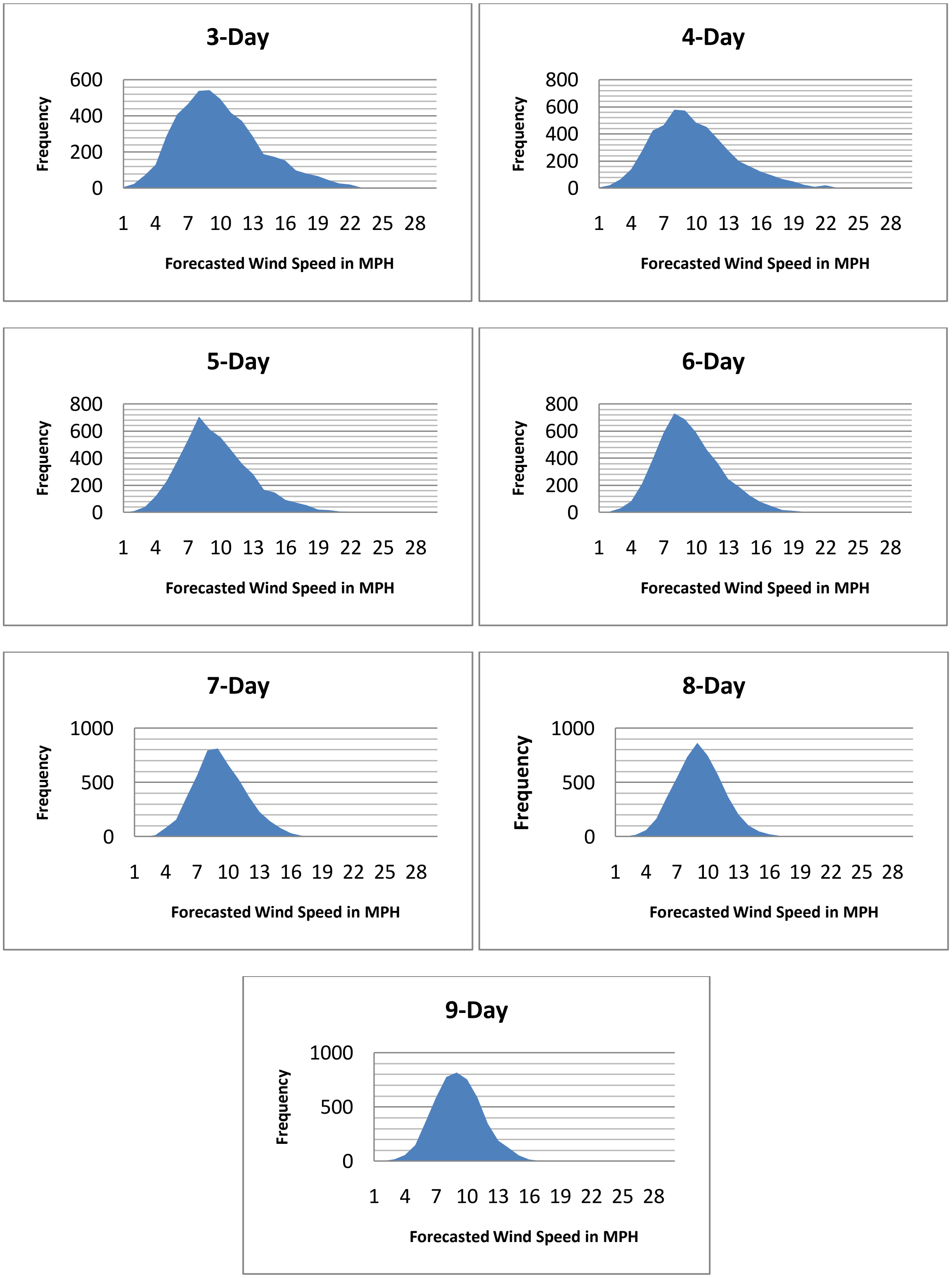


Figure 2b: TWC forecast distribution graphs for 3-9 day lead time for Class 1 locations.

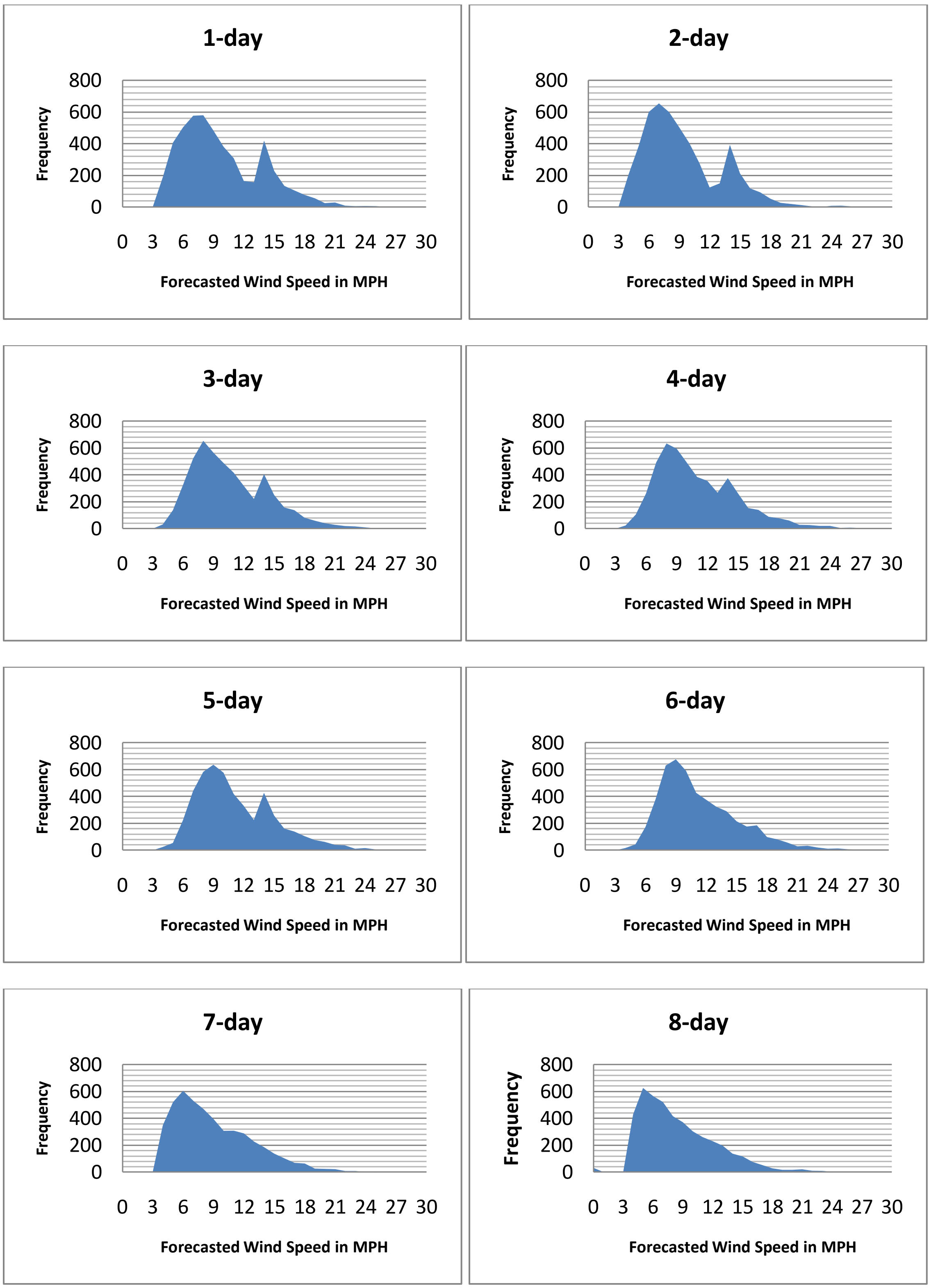


Figure 3a: AWX forecast distribution graphs for 1-8 day lead time for Class 1 locations.

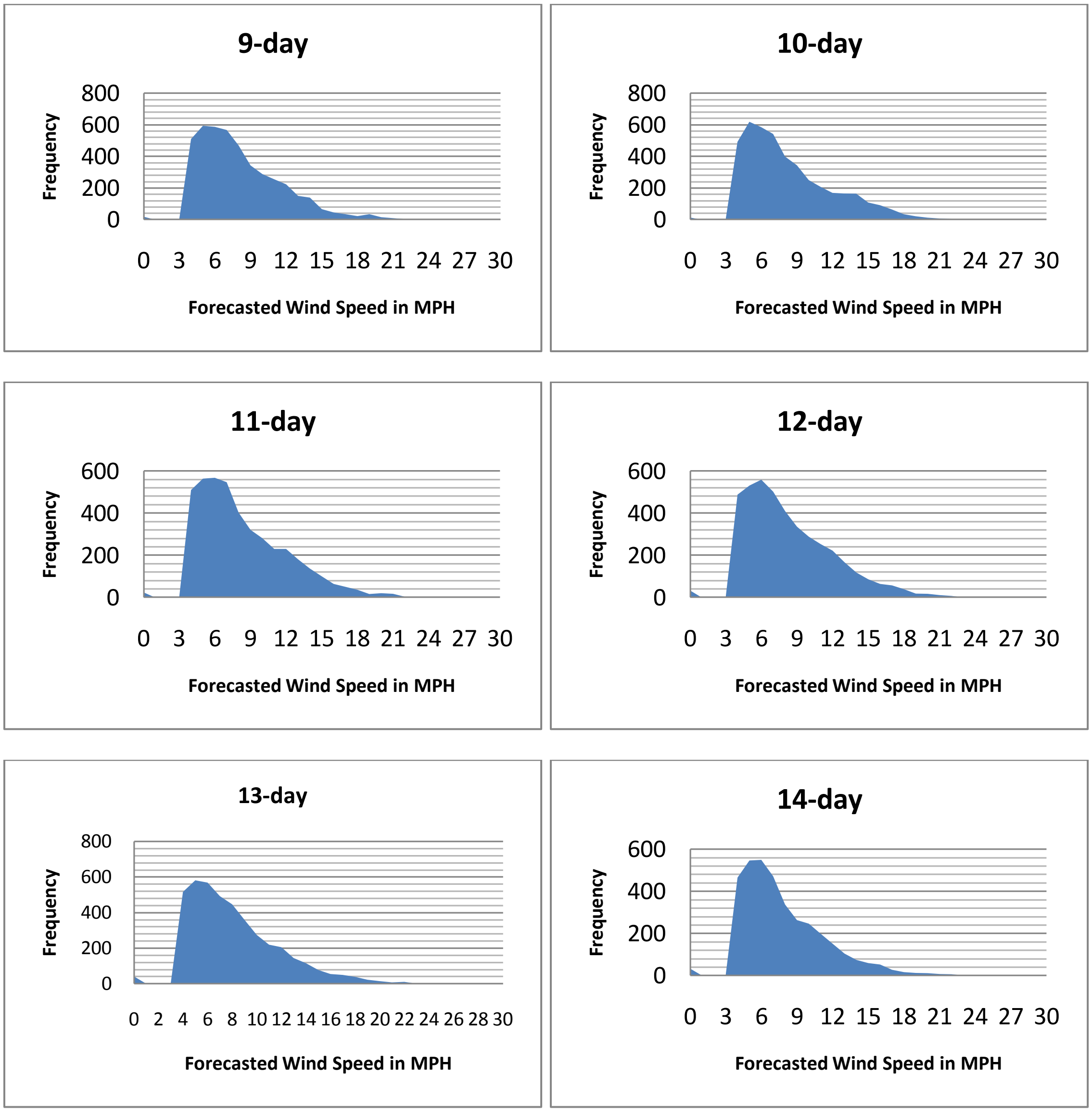


Figure 3b: AWX forecast distribution graphs for 9-14 day lead time for Class 1 locations.

Chapter 5: Forecast Verification

For this analysis section of the report we will be making use of calibration diagrams, MSESS graphs, discrimination value graphs, and illustrating results through several other summary measures. The calibration diagrams include a line at $45°$ that identifies the wind speed forecasts that are perfectly calibrated. This means that given a forecast value, the expected value of the observations is equal to the forecast value. The line of no resolution is a horizontal line through the average observed wind speed, and represents the case that the observations are independent of the forecasts. For clarity within the graphs, the line of no resolution is omitted from the smaller versions of the calibration diagrams; however, the average observed wind speed values located in Tables 2 and 3 can be used to ascertain the location of the line of no resolution for the diagrams. For the purpose of this analysis, forecast values made less than 30 times were excluded in order to obtain a satisfactory estimate of the standard deviation of the forecasts.

We identified a confidence interval around the line of perfect calibration based on the number of forecasts, which allows us to gain a more accurate understanding of which forecasts are well calibrated and which are not. We establish a 99% confidence interval, for which there would be a 1% chance that a forecast-observation pair would be outside the interval. If the expected wind speed was truly *f*, then there would be a 99% chance that the average observed speed would be within

$$f \pm \Phi^{-1}(0.995)\frac{s}{\sqrt{n}}, \tag{9}$$

where Φ^{-1} is the inverse of the standard normal cumulative distribution [$\Phi^{-1}(0.995) = 2.576$] and *n* is the number of forecasts. This interval forms an envelope around the line of perfect calibration, with a width determined by Eq. (9). If a forecast-observation pair lies outside of this interval, then it is not well calibrated. Within the calibration diagrams, forecast-observation pairs that are well calibrated are represented with a black circle, while the pairs that are not well calibrated are symbolized by an open circle.

From Figure 4 we can see that the forecasts of 3, 4, and 5 mph are well calibrated, while the rest are not. While these values are not well calibrated, they are still close to the line of perfect calibration and demonstrate positive skill. Additionally, it can be seen that each pair that is not well calibrated is located below the line of perfect calibration. This illustrates how TWC tends to over-forecast in their 1 and 2 day forecasts. Figure 5 presents the calibration diagrams for 2-9 day lead times. The results of the 2-day forecasts can be seen to be very similar to the 1-day forecasts. For each of the remaining lead times in the forecasting period, TWC appear to under-forecast values on average. Additionally, these diagrams illustrate how the range of values forecast by TWC begins to shrink as the lead time increases. While the majority of forecasted values are not well calibrated, they demonstrate positive skill even 9 days out.

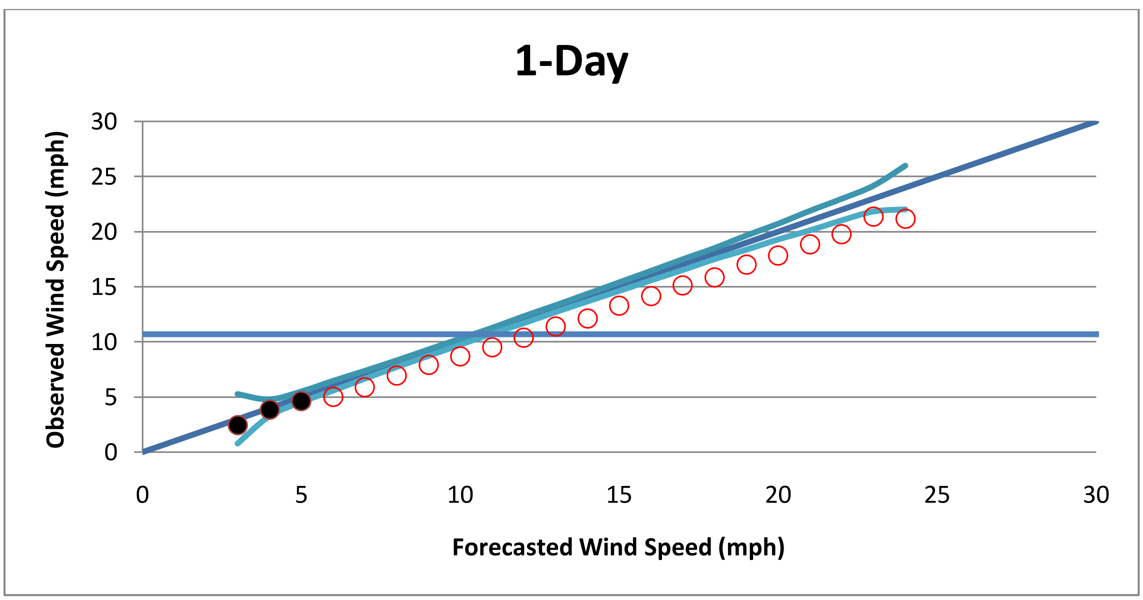


Figure 4a: Calibration diagram for TWC's 1-day wind speed forecasts for Class 1 locations.

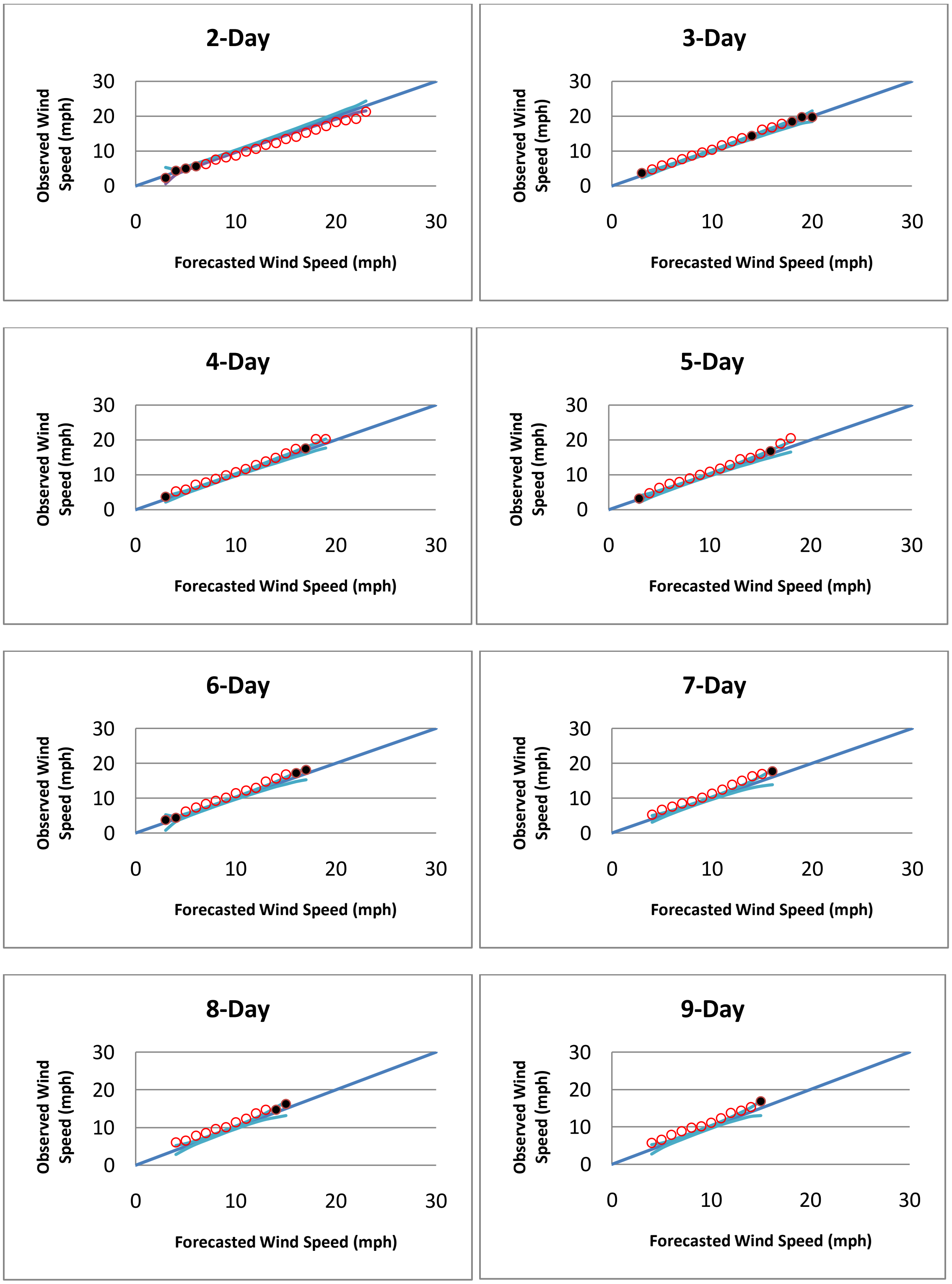


Figure 4b: Calibration diagrams for TWC's 2-9 day lead times for Class 1 locations.

Figure 6 presents the calibration diagrams for AWX's forecasts for Class 1 locations. Unlike TWC's 1 and 2 day forecasts, AWX's tends to under-forecast on average. The forecast-observation pairs that are not well calibrated tend to stay close to the line of perfect calibration and have positive skill until the 7-day lead time. From this point on the forecasts have either no skill or negative skill, and by the end of the forecasting period the forecasts appear to lay on the line of no resolution.

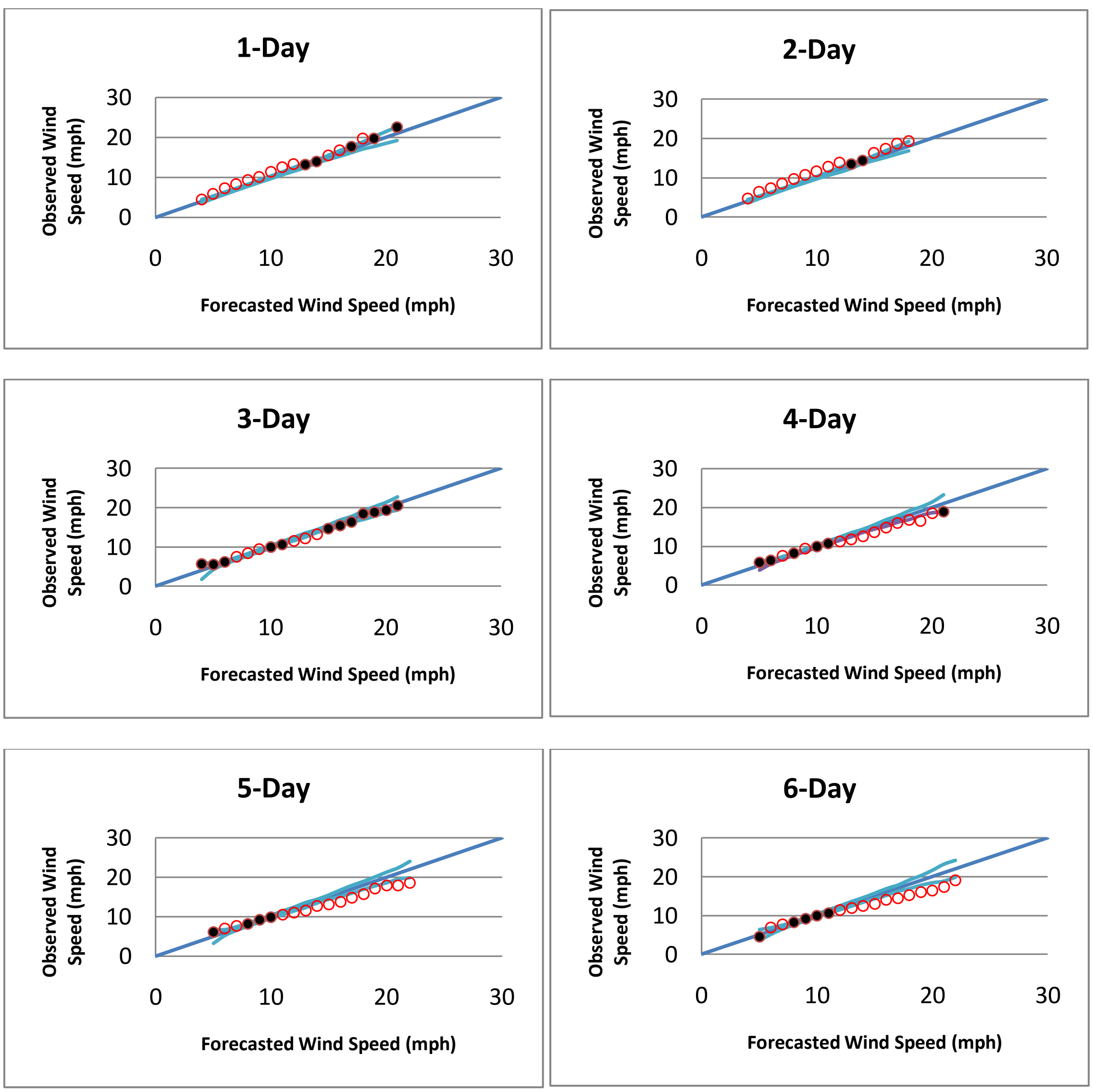


Figure 5a: Calibration diagrams for AWX's 1-6 day lead times for Class 1 locations.

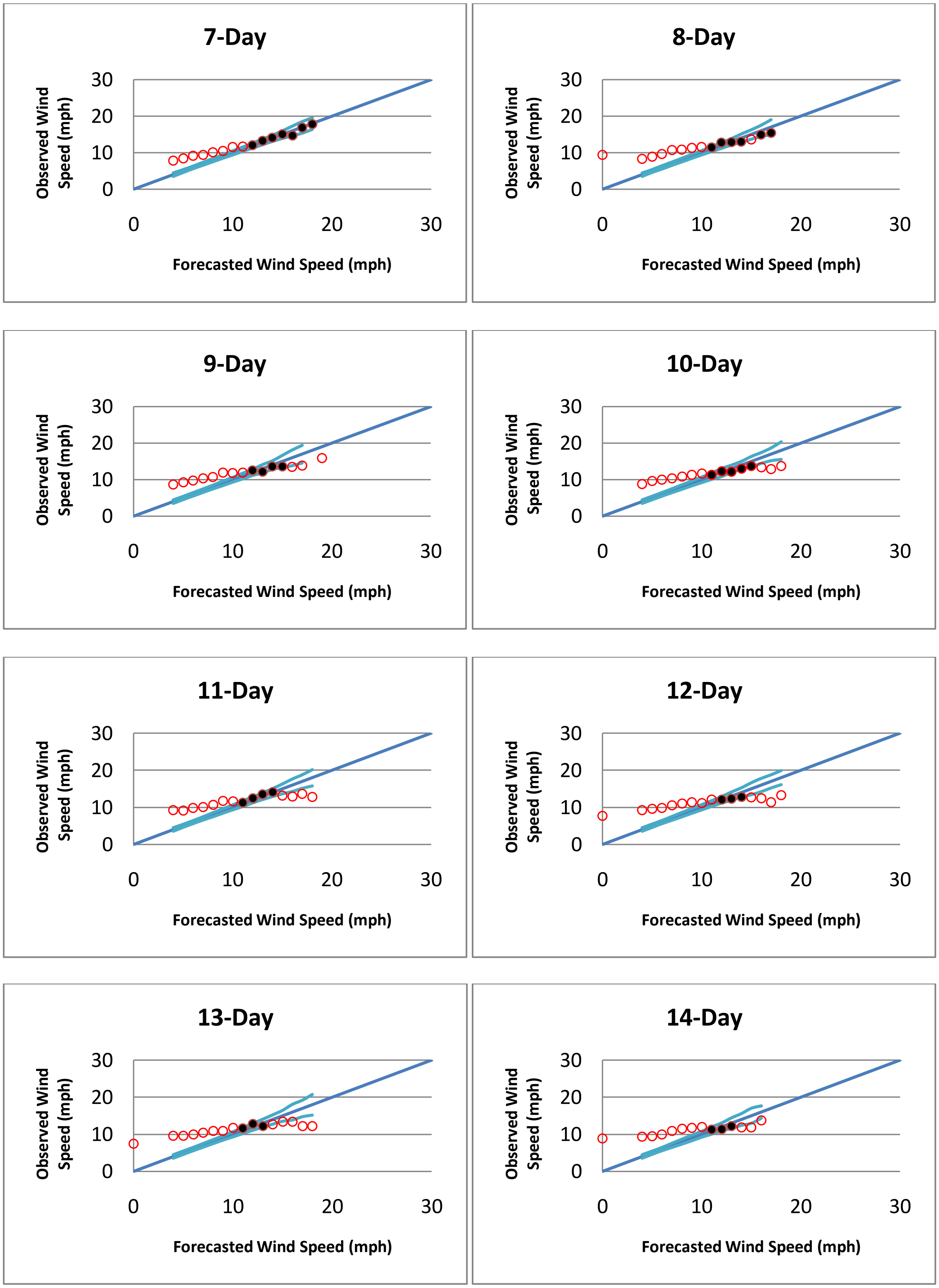


Figure 5b: Calibration diagrams for AWX's 7-14 day lead times for Class 1 locations.

Table 4 presents several summary measures of forecast performance, including the variance of the forecasts. For deterministic forecasts of discrete or continuous variables, sharpness is most simply estimated by the variance of the forecasts. The variance of TWC's forecasts is always less than the variance of the observations, but for the first 2 days of the forecast period it is comparable. After the first 2 days the variance of the forecasts decreases every day until by the 9-day forecast the variance of the forecasts is only between 15 and 25 percent of the size of the variance of observations, depending on the wind power class. This illustrates that while TWC forecasts have positive skill for all lead times, they are not sharp for the larger lead time forecasts.

As was the case for TWC, the variance of AWX's forecasts is always less than the variance of the observations. Initially, the variance of these forecasts tend to be slightly less than TWC's; however, they remain largely unchanged throughout the forecasting period. The variance of the forecasts is always at least 51 to 70 percent of the size of the variance of the observations, depending on the wind power class. This demonstrates that while the calibration and skill of the AWX forecasts deteriorate with time, they remain sharp throughout the forecasting period.

Also presented in Table 4 are the mean-square error and the correlation. The MSE, calculated with respect to mph, ranges from 8.516 to 78.537 for the AWX forecasts and between 8.032 and 34.522 for TWC's forecasts. The correlation between TWC's forecasts and the observations begins at over 0.83 for all 4 wind power classes, and then declines with each subsequent day. Other than for wind power class 4, the correlation remains above 0.4 for all lead times. For AWX the correlation begins at over 0.77 for all 4 wind power classes, before declining more sharply than TWC with lead time. The correlation drops to just over 0.2 for all wind power classes other than 4, which drops down to 0.084.

Table 4a: Summary measures of forecasting performance at varying lead times for TWC for Class 1-3 locations.

TWC Class 1

	Variance			
Lead Time (days)	Forecasts	Observations	MSE	Correlation
1	21.193	23.450	8.548	0.860
2	21.167	23.508	8.932	0.831
3	16.077	23.536	8.621	0.808
4	14.493	23.576	10.743	0.758
5	11.315	23.506	13.232	0.694
6	8.787	23.508	15.630	0.634
7	6.394	23.624	18.263	0.558
8	5.730	23.492	20.134	0.481
9	5.445	23.510	20.821	0.453

TWC Class 2

	Variance			
Lead Time (days)	Forecasts	Observations	MSE	Correlation
1	21.354	25.654	8.032	0.848
2	21.690	25.719	8.814	0.824
3	17.182	25.746	10.969	0.781
4	15.457	25.769	13.423	0.727
5	12.107	25.766	16.551	0.652
6	9.855	25.819	19.283	0.580
7	7.251	25.872	21.971	0.507
8	6.523	25.758	23.490	0.444
9	6.331	25.529	24.321	0.404

TWC Class 3

	Variance			
Lead Time (days)	Forecasts	Observations	MSE	Correlation
1	22.052	29.219	9.559	0.831
2	21.143	29.322	10.370	0.809
3	19.210	29.169	12.302	0.778
4	16.668	29.370	14.447	0.742
5	13.822	29.360	18.308	0.655
6	11.273	29.199	20.773	0.595
7	8.355	29.270	23.648	0.524
8	7.649	29.192	24.560	0.489
9	7.442	29.012	25.687	0.439

Table 4b: Summary measures of forecasting performance at varying lead times for TWC for Class 4 locations and AWX for Class 1 and 2 locations.

TWC Class 4

	Variance			
Lead Time (days)	Forecasts	Observations	MSE	Correlation
1	31.075	35.548	10.428	0.848
2	31.127	35.761	14.422	0.787
3	24.780	35.752	17.252	0.736
4	20.407	35.639	21.497	0.656
5	14.162	35.554	26.146	0.551
6	10.481	35.190	29.040	0.473
7	6.509	34.787	32.412	0.354
8	5.798	35.456	33.099	0.353
9	5.436	34.793	34.522	0.277

AWX Class 1

	Variance			
Lead Time (days)	Forecasts	Observations	MSE	Correlation
1	16.704	22.467	8.516	0.819
2	15.006	22.842	11.190	0.772
3	15.689	23.398	11.979	0.707
4	16.833	23.457	14.756	0.647
5	16.204	23.435	16.496	0.603
6	15.860	23.513	18.317	0.556
7	15.063	23.686	23.601	0.485
8	14.409	23.622	30.425	0.361
9	12.926	23.927	32.683	0.314
10	14.324	23.992	34.316	0.274
11	14.239	23.328	33.047	0.285
12	14.081	23.502	34.737	0.240
13	13.343	23.434	35.063	0.247
14	11.911	23.134	36.376	0.217

AWX Class 2

	Variance			
Lead Time (days)	Forecasts	Observations	MSE	Correlation
1	19.507	25.604	8.918	0.836
2	16.801	25.685	11.887	0.781
3	17.714	25.838	14.492	0.680
4	17.990	25.888	17.071	0.624
5	16.614	25.820	19.881	0.553
6	17.711	25.847	21.947	0.524
7	19.130	25.892	28.306	0.456
8	18.637	25.901	34.929	0.369
9	16.794	25.598	38.793	0.280
10	20.235	25.941	38.888	0.298
11	20.766	25.600	39.467	0.288
12	19.518	25.730	42.015	0.233
13	18.588	25.796	41.698	0.245
14	17.080	25.479	40.718	0.260

Table 4c: Summary measures of forecasting performance at varying lead times for AWX for Class 3 and 4 locations.

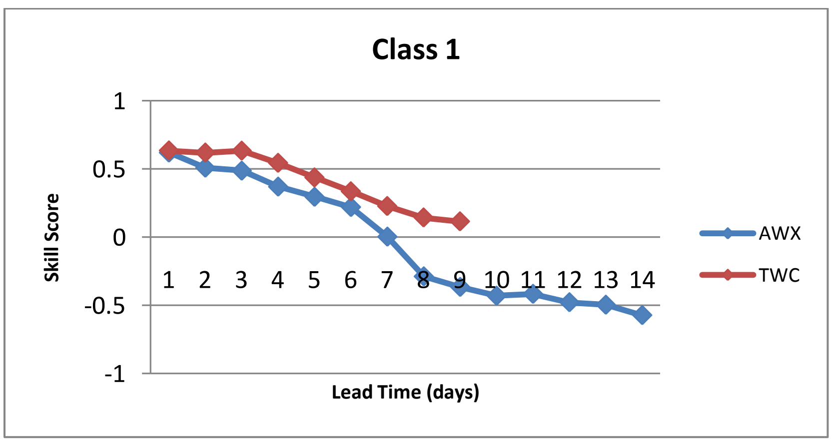
AWX Class 3

	Variance			
Lead Time (days)	Forecasts	Observations	MSE	Correlation
1	22.582	29.182	11.105	0.814
2	20.445	29.240	14.421	0.754
3	21.657	29.146	17.392	0.665
4	21.636	29.377	20.672	0.602
5	20.427	29.326	22.914	0.550
6	21.441	29.244	25.155	0.517
7	19.856	29.202	32.965	0.436
8	21.978	29.239	42.228	0.334
9	20.728	29.009	46.018	0.273
10	26.823	28.924	46.352	0.283
11	25.460	28.879	45.932	0.271
12	25.459	28.937	49.309	0.217
13	25.559	28.738	51.173	0.221
14	24.158	28.872	51.027	0.211

AWX Class 4

	Variance			
Lead Time (days)	Forecasts	Observations	MSE	Correlation
1	28.469	35.540	15.271	0.777
2	26.195	35.576	19.120	0.722
3	33.344	35.666	25.011	0.640
4	27.486	35.749	28.029	0.561
5	30.416	35.473	38.956	0.414
6	27.062	35.383	36.749	0.434
7	24.693	35.341	52.349	0.297
8	25.576	35.540	66.632	0.173
9	24.010	34.309	67.792	0.126
10	28.969	34.880	70.886	0.134
11	27.597	34.694	67.527	0.179
12	25.738	34.638	71.513	0.118
13	24.468	34.591	78.109	0.097
14	25.988	35.014	78.537	0.084

Figure 6 presents the mean-square error skill score graphs for both providers. The graphs for TWC show that their forecast skill score remained positive for all lead times and all wind power classes. The skill scores for TWC range between approximately 0.71 and 0.01, depending on lead time and wind power class. The decline in the variance of TWC's forecasts mirrors the decline in the skill scores of the forecasts. This shows that for larger lead times, TWC is becoming more and more conservative with their forecasting strategy. The end result of this strategy is that the forecasts have value slightly higher than the climatological average, but have minimal sharpness. The graphs for AWX show that their forecasts decrease in skill faster than TWC's, and eventually attain negative values. These negative values mean that the forecasts are worse than the climatological average forecast. With the exception of class 1 locations, all of AWX forecasts attain negative skill by the 7-day lead time. The overall range of skill scores for the AWX forecasts are between 0.65 and -1.26, depending on lead time and wind power class. These results combined with the discrimination results show that AWX continues to have sharp forecasts throughout the forecasting period at the cost of overall forecast skill.



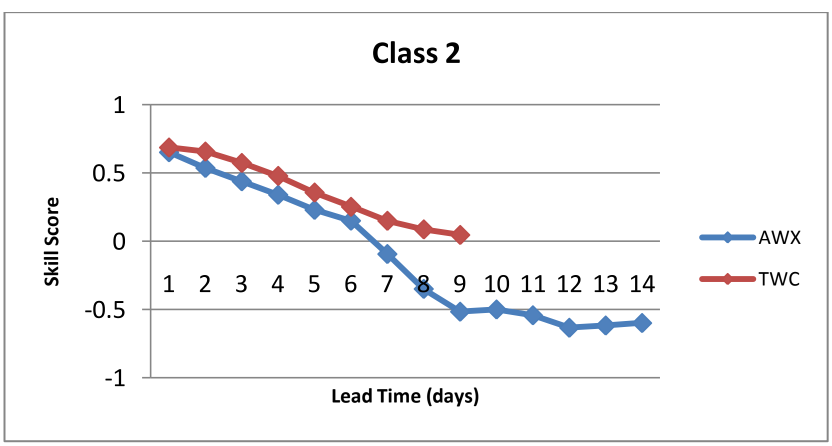
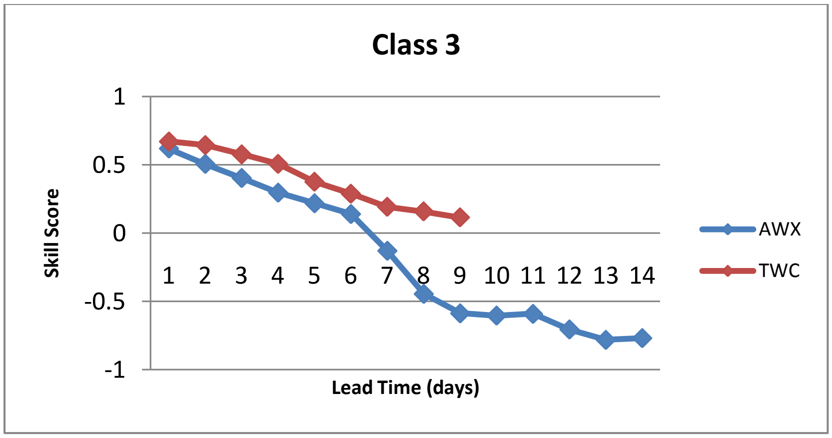


Figure 6a: Mean-square error skill score graphs for wind power classes 1-2 for both providers.



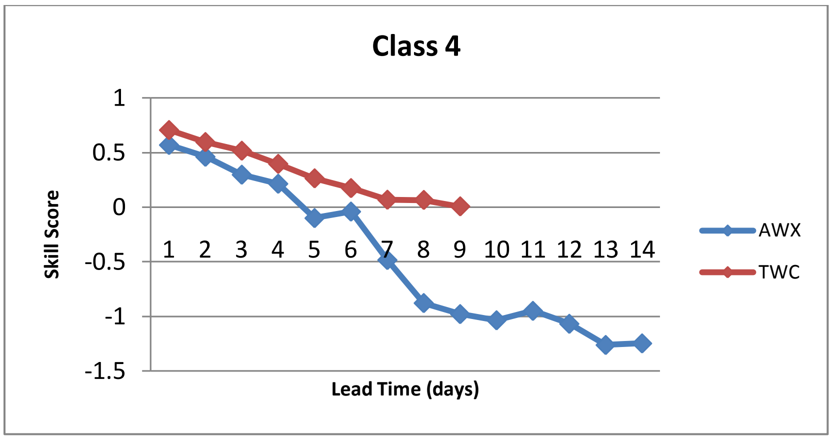
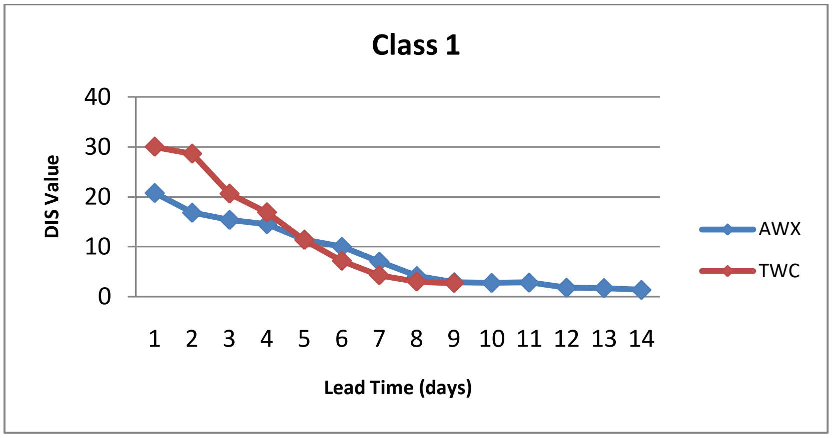
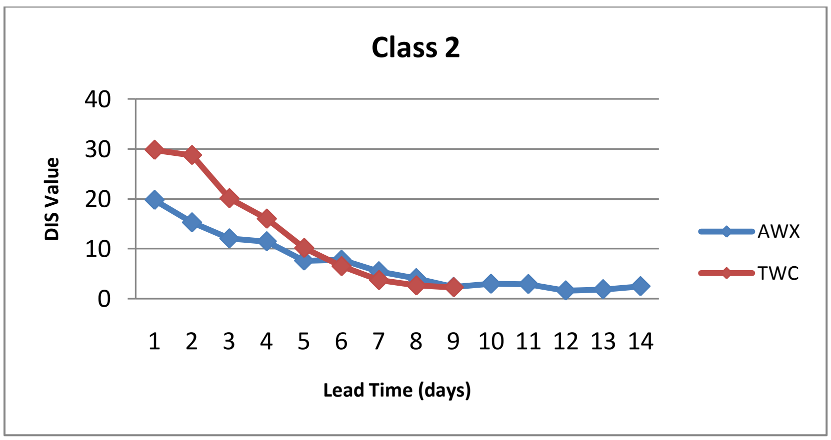


Figure 6b: Mean-square error skill score graphs for wind power classes 3-4 for both providers.

Lastly, we examine the discrimination scores based on the likelihood-base-rate factorization. Figure 7 presents the discrimination value scores for both providers for all lead times. From these graphs we can see that the forecasts for TWC have a higher DIS score than those for AWX for the first 4 days of their forecasts in all wind power classes. After this point, it appears AWX discrimination score is either equal to or higher than that of TWC. The DIS scores for 1-day TWC forecasts range from approximately 30 in Class 1 locations, to just over 16 in Class 4 locations. For 1-Day AWX forecasts, the DIS scores range from approximately 21 in Class 1 locations, to just over 10 in Class 4 locations. This illustrates that both TWC and AWX have their worst performances in the Class 4 locations, and that their DIS scores are significantly lower for those areas.





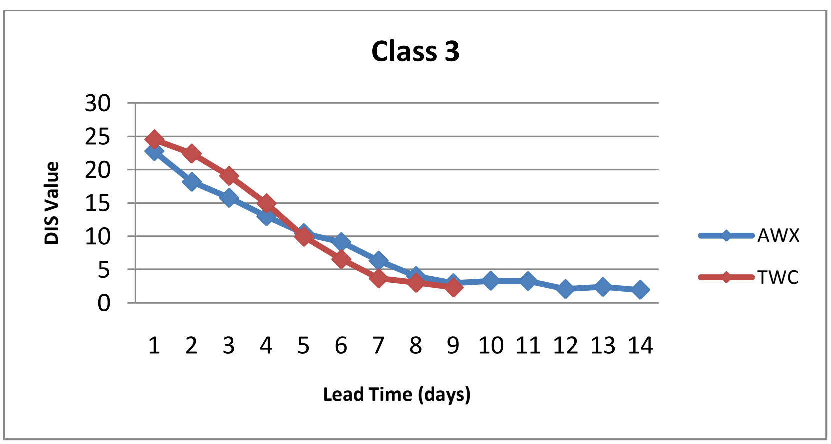


Figure 7a: Discrimination value graphs for both providers for Classes 1-3.

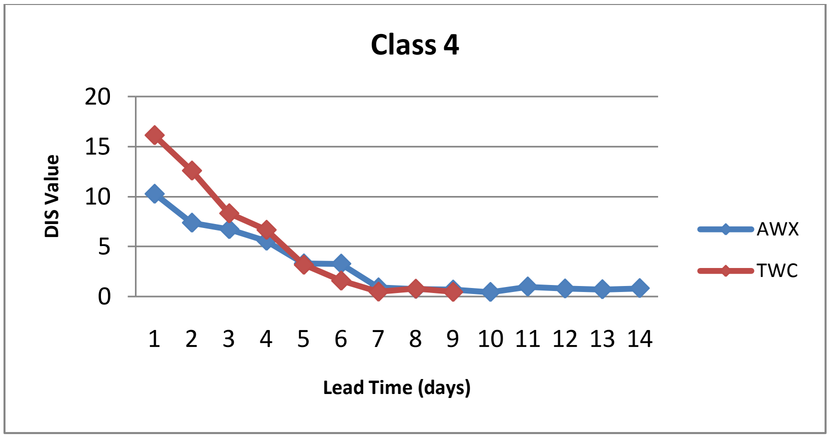


Figure 7b: Discrimination value graph for both providers for Class 4.

Chapter 6: Conclusions

The preceding analysis has provided us with interesting insights into the forecasting systems of both TWC and AWX. Looking at TWC first, we see that their forecasts exhibit positive skill for all lead times and all wind power classes. While the majority of TWC's forecasts are not well calibrated, all of their forecasted values are close to the line of perfect calibration and demonstrate positive skill that decreases steadily with lead time. TWC forecasts are sharp for early forecasts, but their sharpness decreases consistently with lead time. This forecasting strategy allows for TWC to maintain a positive level of skill and few poorly calibrated values. Lastly, TWC tends to over-forecast wind speed for lead times less than 3 days, especially for the lower wind power classes. Increasing the accuracy of these short term forecasts is one possible area that TWC could improve its forecasting system.

AWX forecasts exhibit positive skill for lead times less than 7 days for all wind power classes. After the first 7 days of the forecast, their skill becomes negative for all wind power classes. AWX forecasts begin to diverge from the line of perfect calibration at the 7-day lead time, and begin to show negative skill until by the end of the forecasting period the majority of the values forecast lie on the line of no resolution. AWX forecasts can be seen to be sharp throughout the forecasting period, and from the data it appears that AWX can successfully create sharp forecasts with positive skill up to 6-day lead times. However, AWX forecasts for larger lead times appear to be too sharp, leading to negative skill scores. AWX would likely benefit from sacrificing some sharpness in longer term forecasts in exchange for increased skill scores. Additionally, AWX tends to under-forecast wind speed for lead times less than 3 days and could benefit from attempting to increase the accuracy of these short term forecasts.

MSE skill scores and discrimination score values tended to be worse within the class 4 regions for both providers. This illustrates that improving forecast accuracy in regions with higher average wind speeds should be a priority for both companies in the future.

Appendix

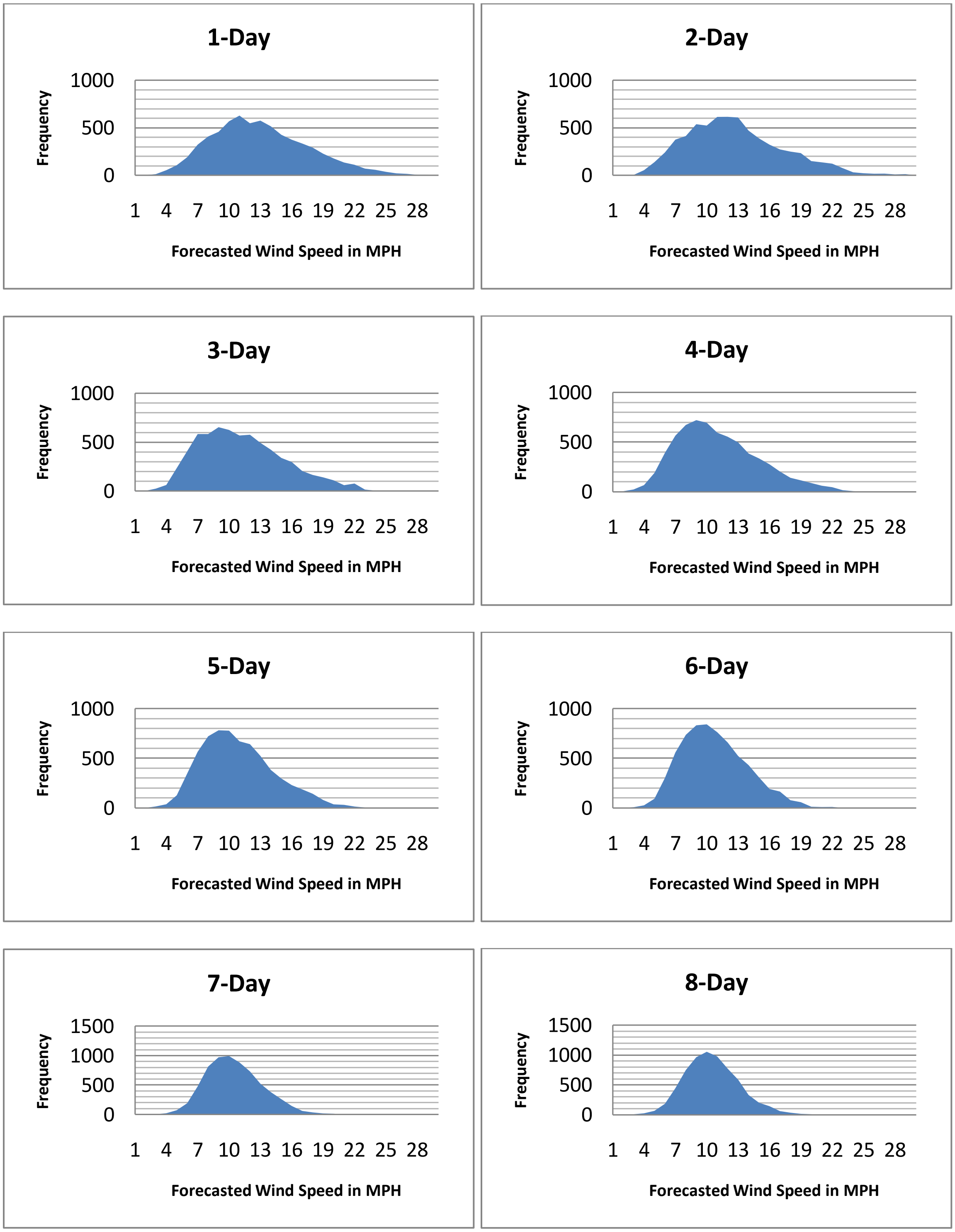


Figure A1a: TWC forecast distribution graphs for 1-8 day lead time for Class 2 locations.

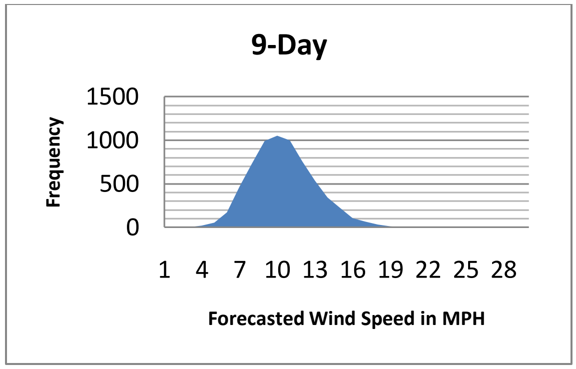


Figure A1b: TWC forecast distribution graphs for 9 day lead time for Class 2 locations.

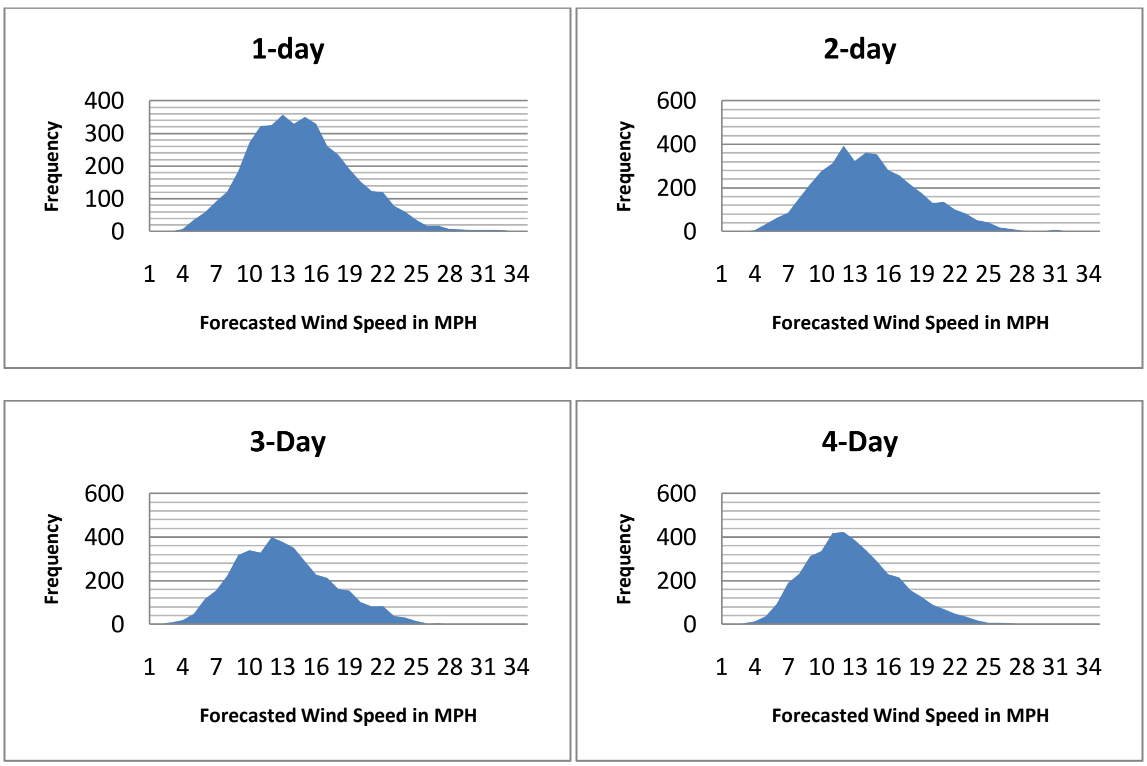


Figure A2a: TWC forecast distribution graphs for 1-4 day lead time for Class 3 locations.

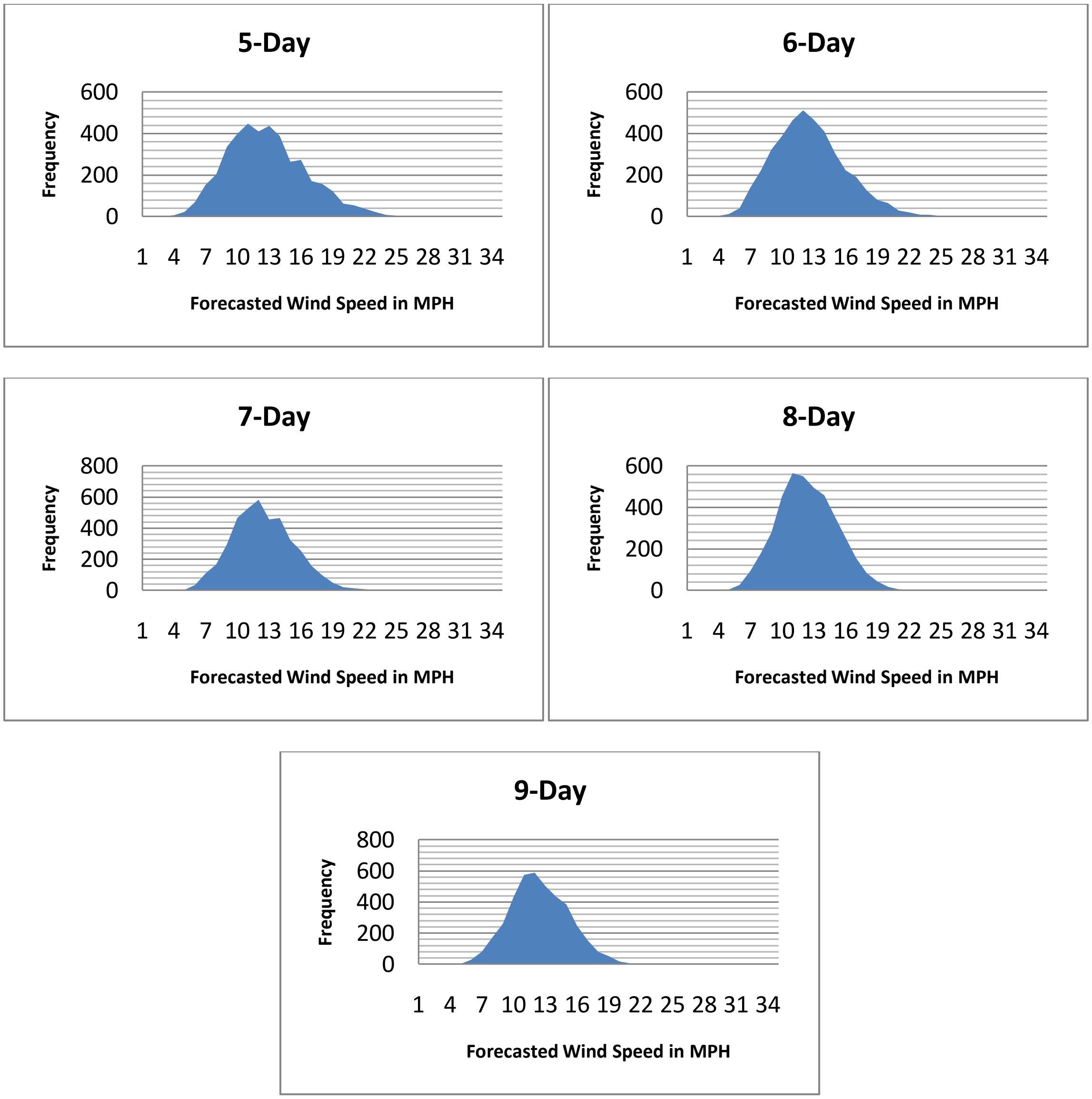


Figure A2b: TWC forecast distribution graphs for 5-9 day lead time for Class 3 locations.

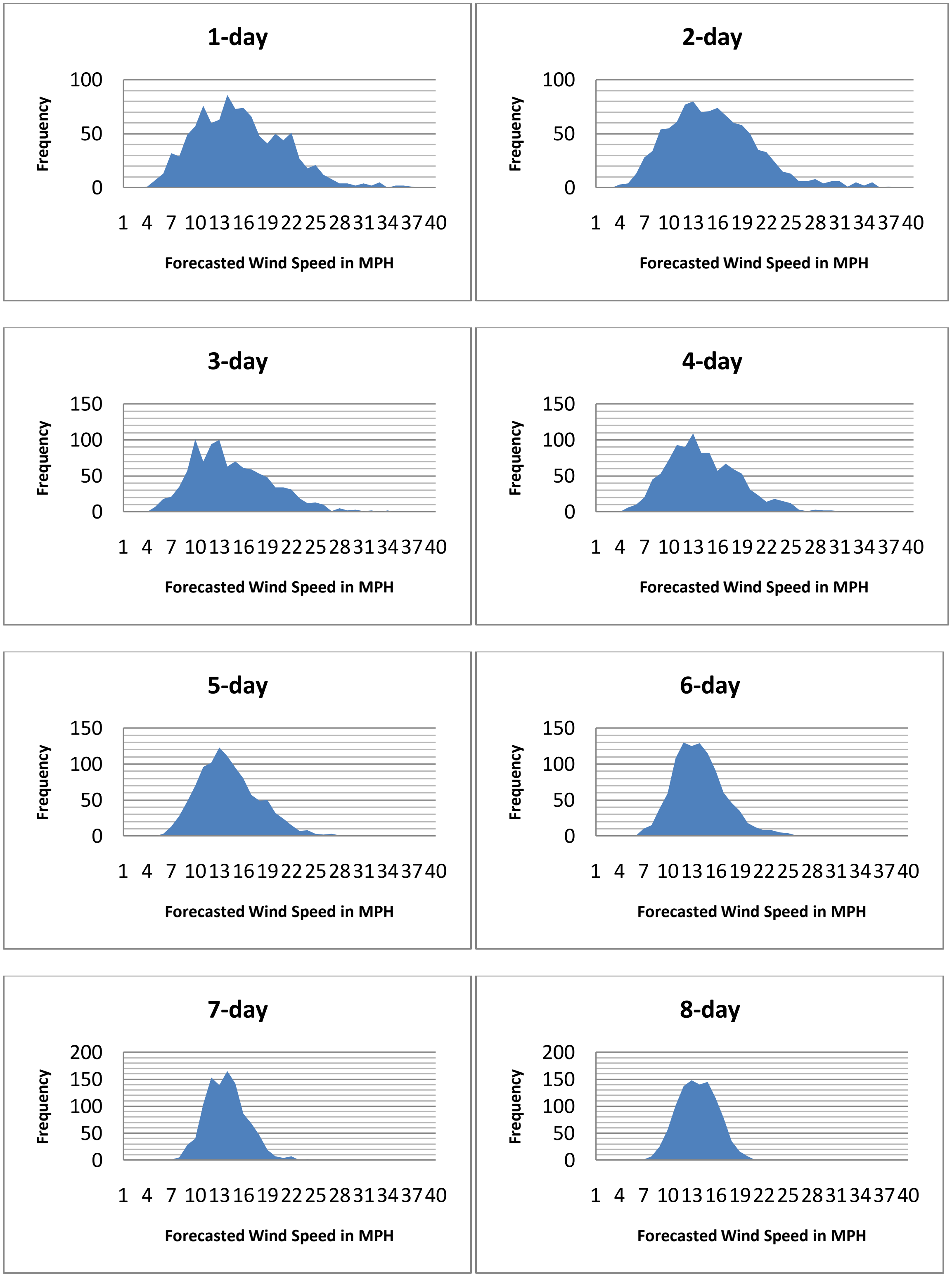


Figure A3a: TWC forecast distribution graphs for 1-8 day lead time for Class 4 locations.

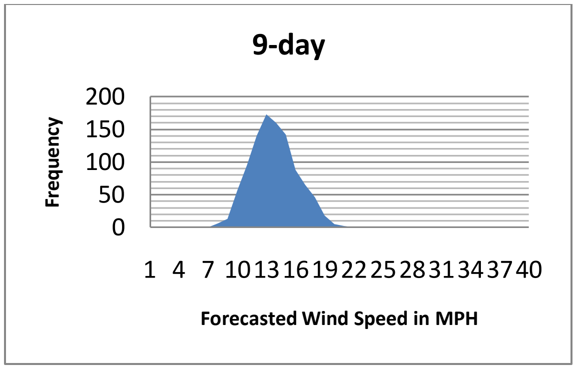


Figure A3b: TWC forecast distribution graphs for 9 day lead time for Class 4 locations.

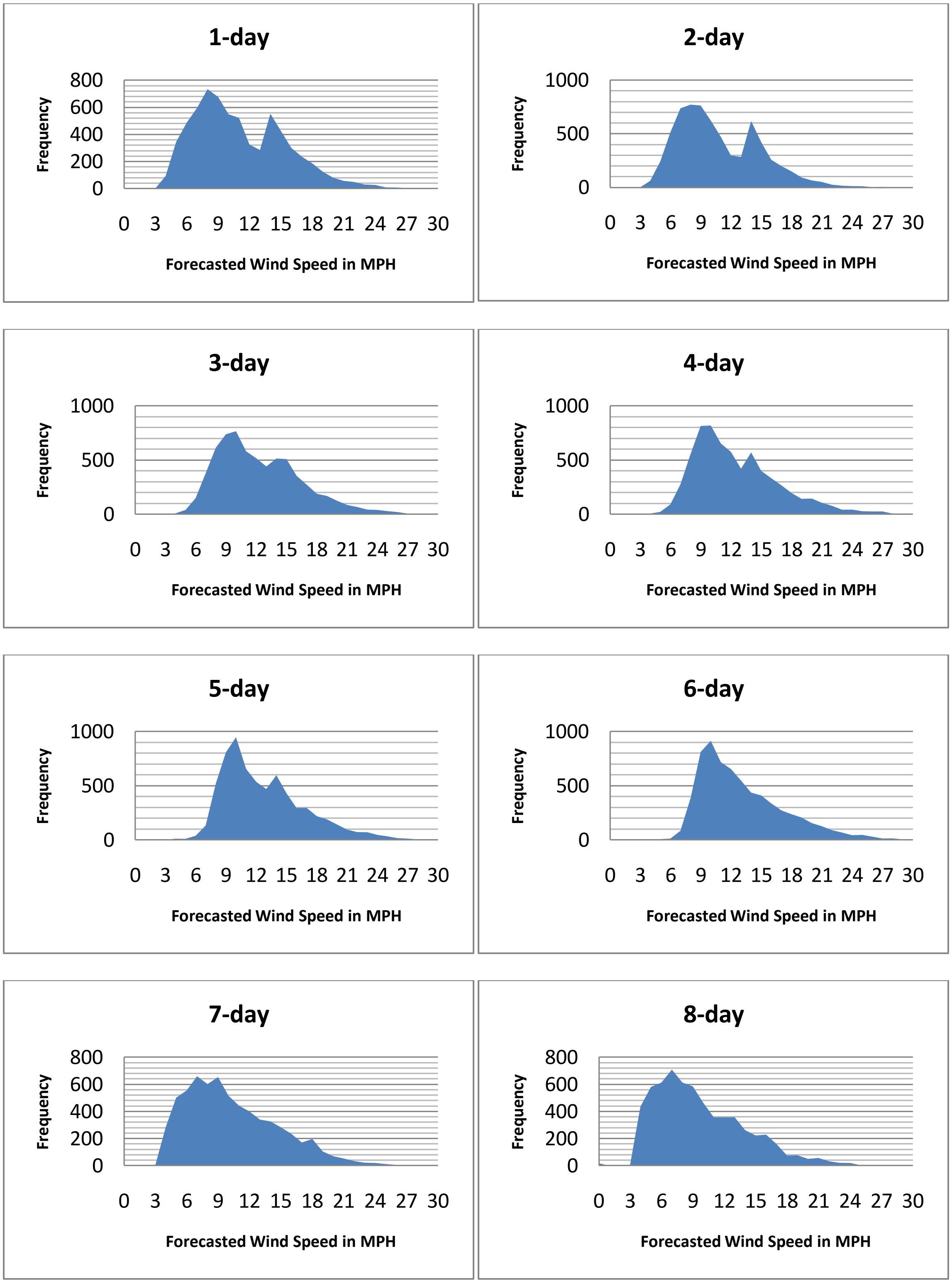


Figure A4a: AWX forecast distribution graphs for 1-8 day lead time for Class 2 locations.

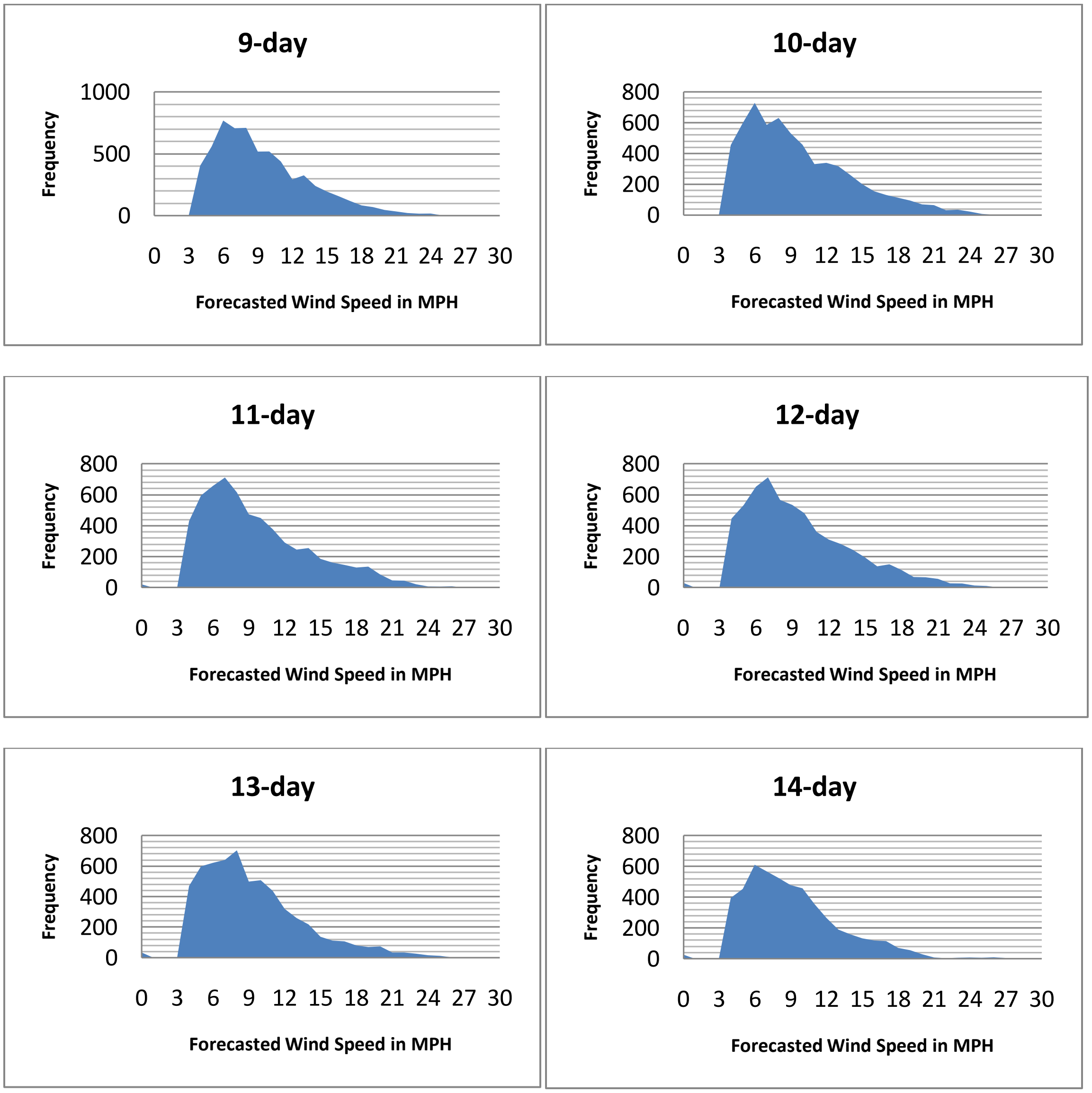


Figure A4b: AWX forecast distribution graphs for 9-14 day lead time for Class 2 locations.

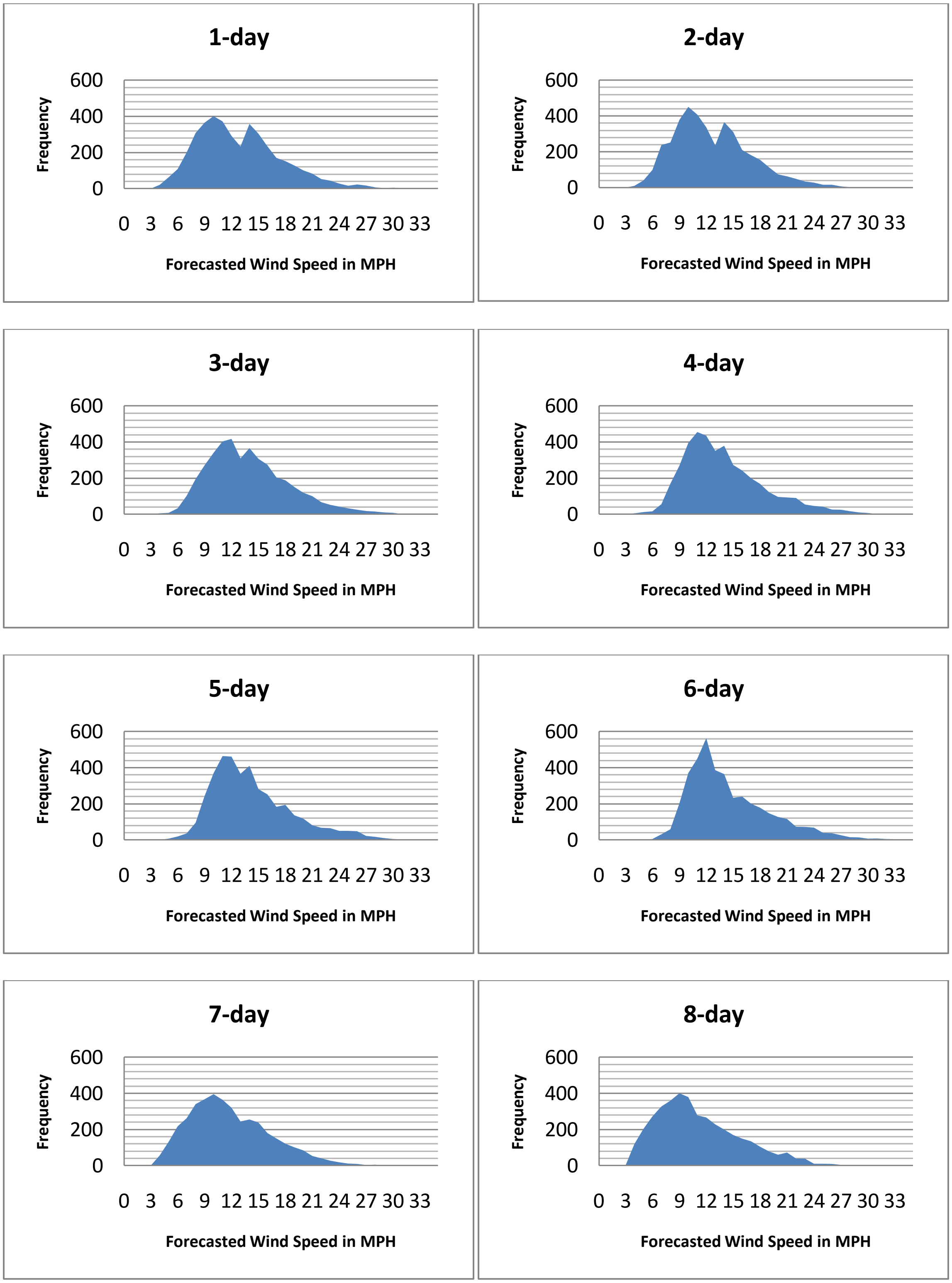


Figure A5a: AWX forecast distribution graphs for 1-8 day lead time for Class 3 locations.

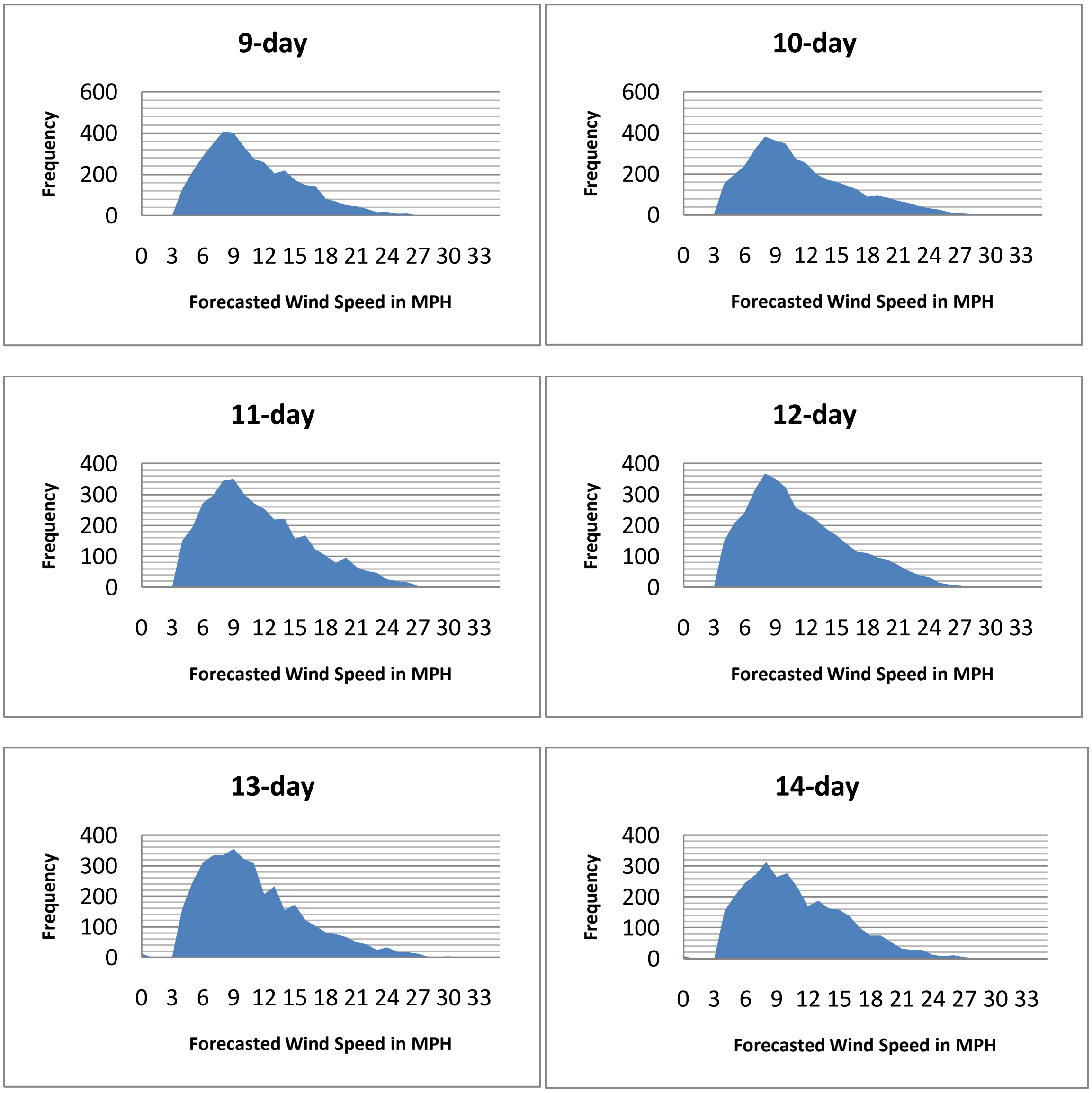


Figure A5b: AWX forecast distribution graphs for 9-14 day lead time for Class 3 locations.

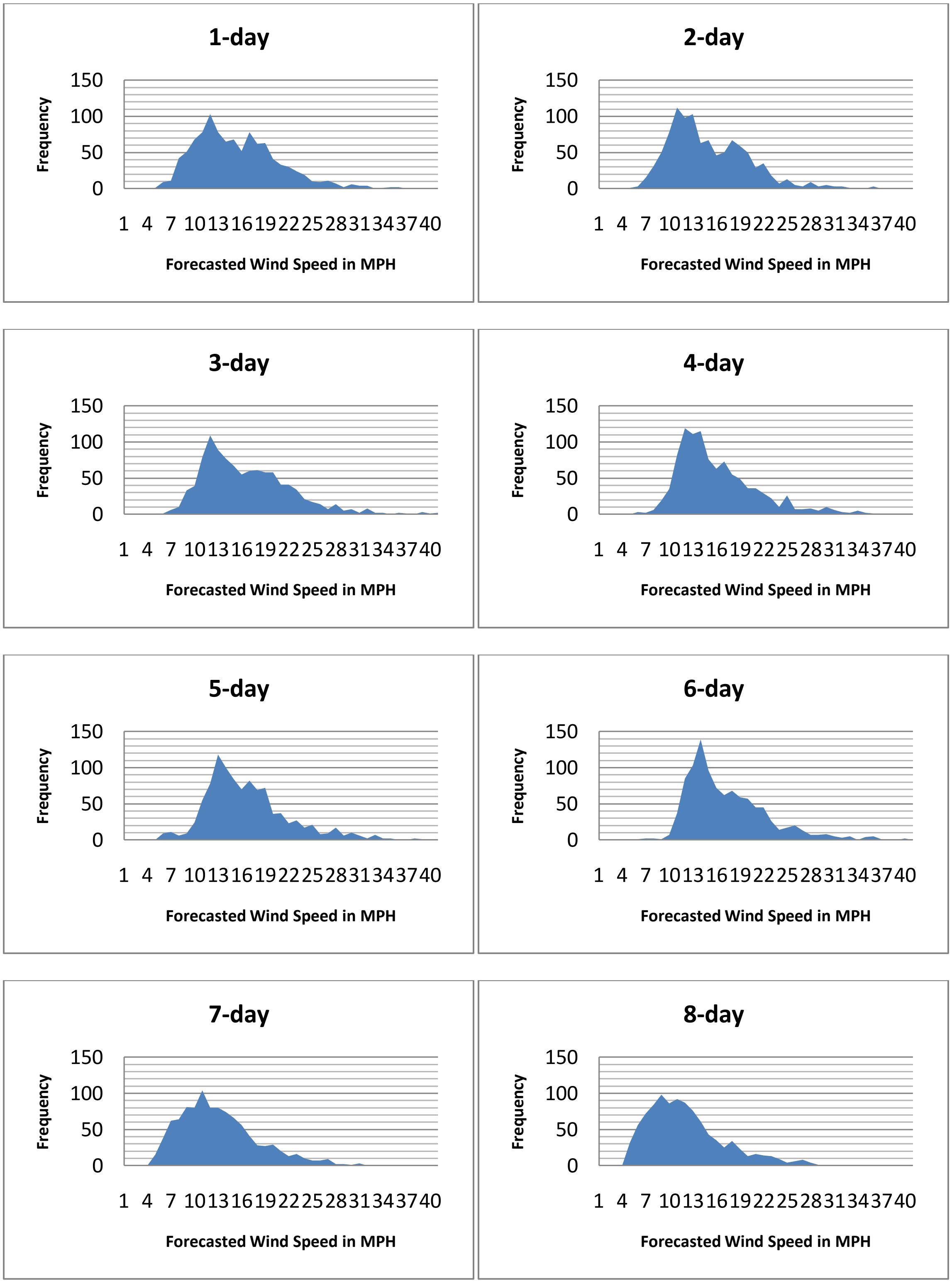


Figure A6a: AWX forecast distribution graphs for 1-8 day lead time for Class 4 locations.

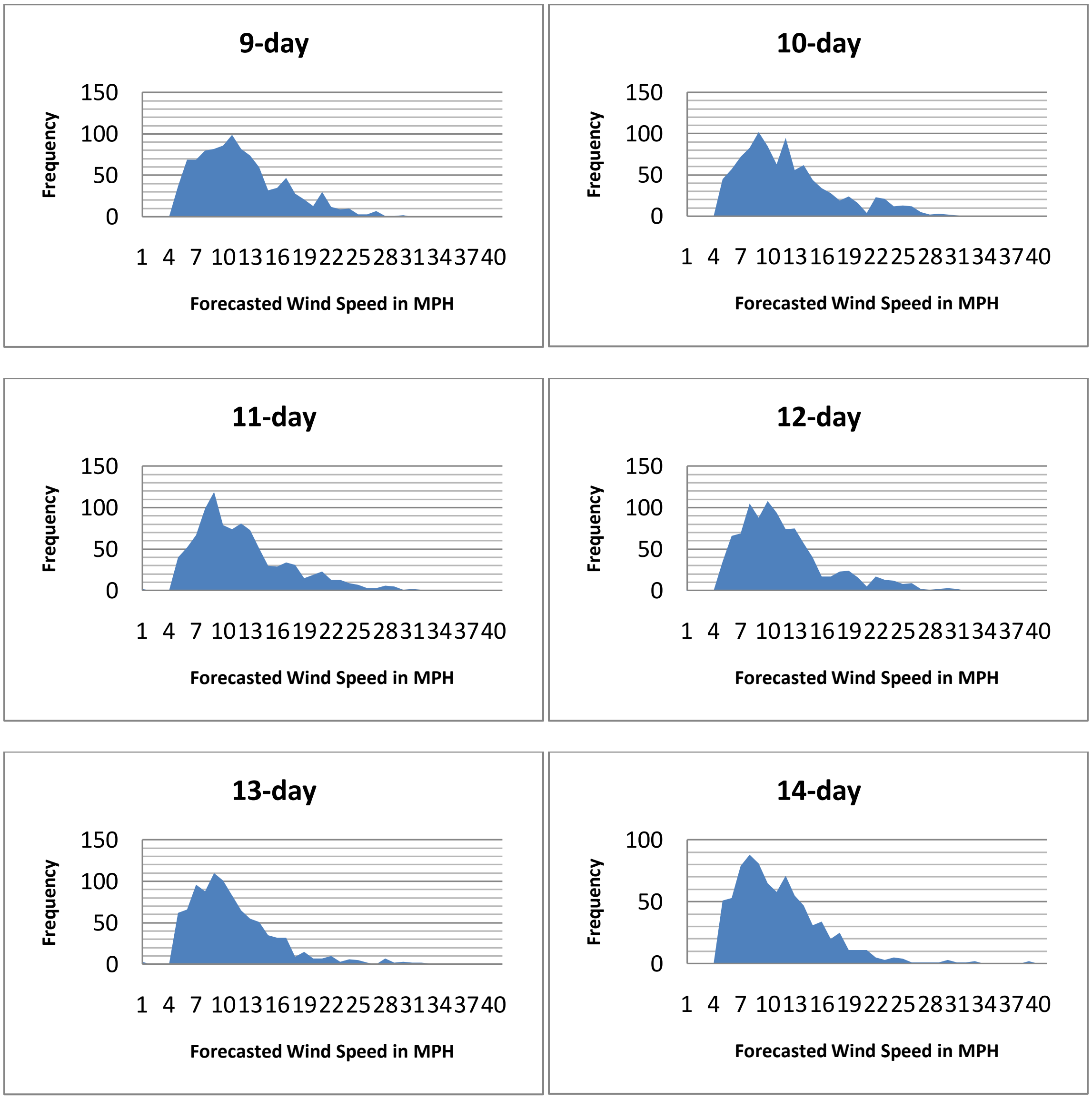


Figure A6b: AWX forecast distribution graphs for 9-14 day lead time for Class 4 locations.

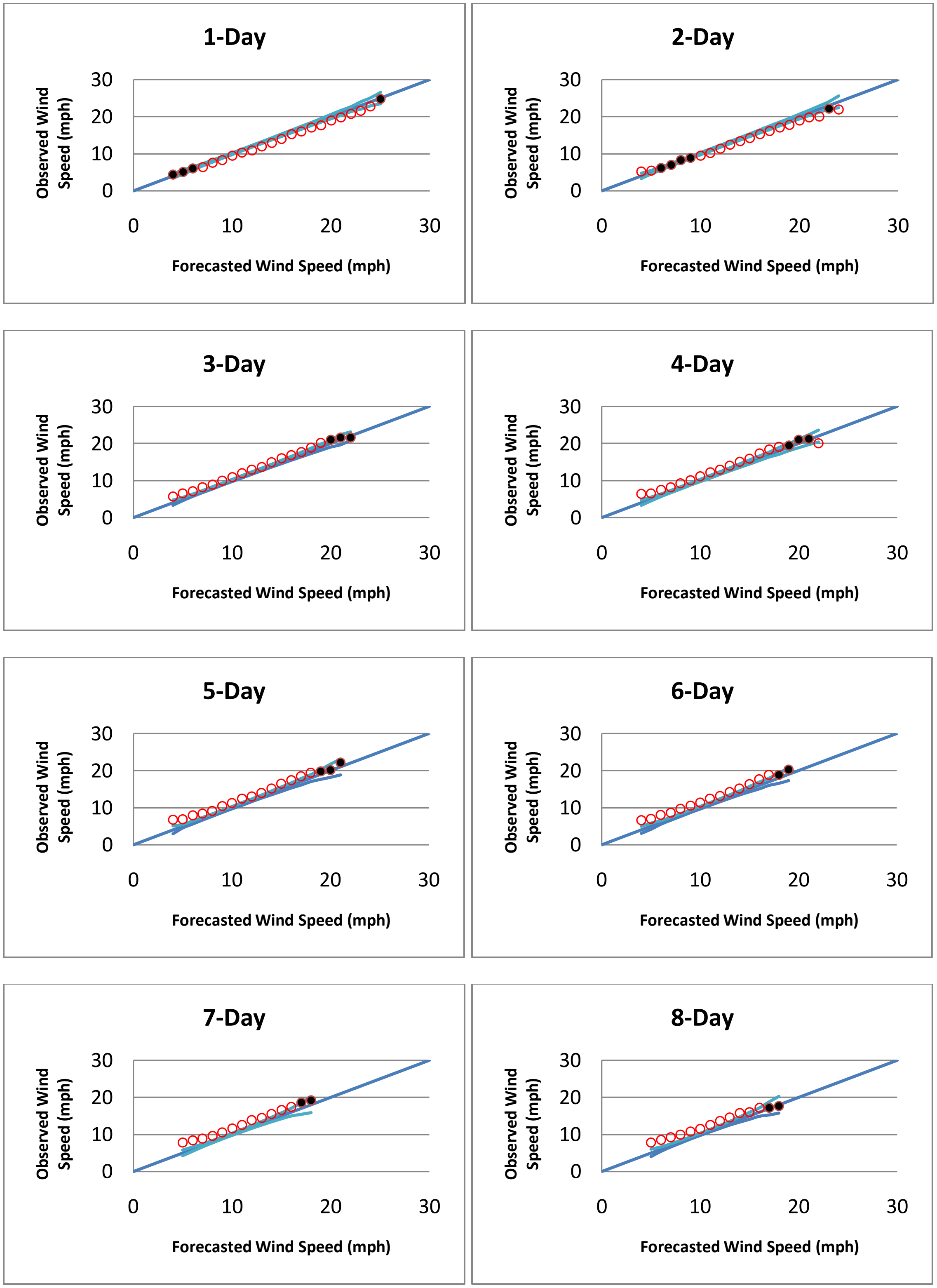


Figure A7a: Calibration Diagrams for TWC's 1-8 day lead times wind speed forecasts for Class 2 locations.

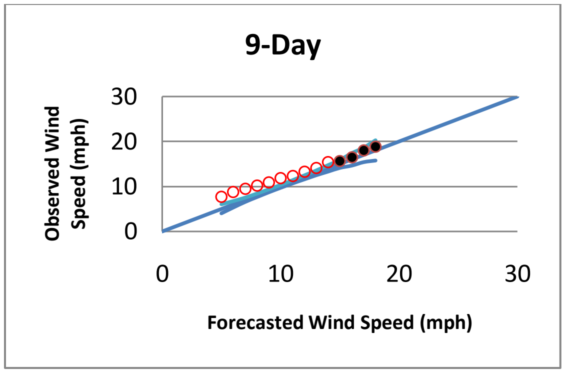


Figure A7b: Calibration Diagrams for TWC's 9 day lead times wind speed forecasts for Class 2 locations.

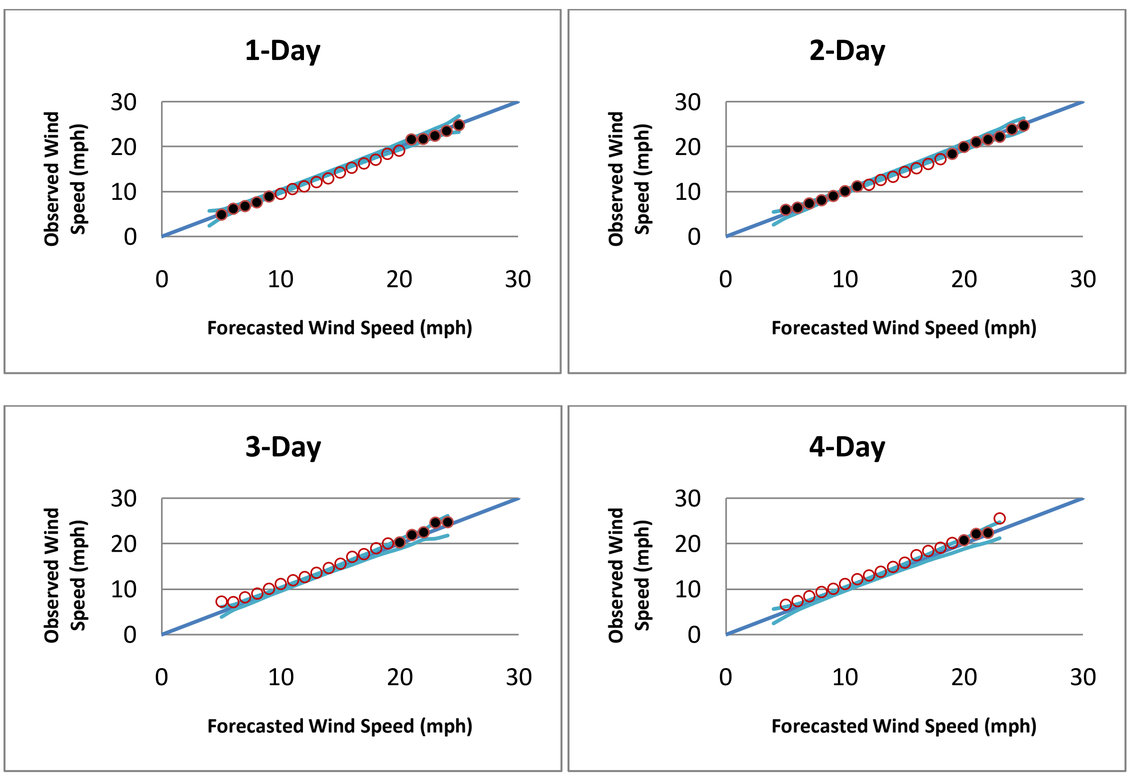


Figure A8a: Calibration Diagrams for TWC's 1-4 day lead times wind speed forecasts for Class 3 locations.

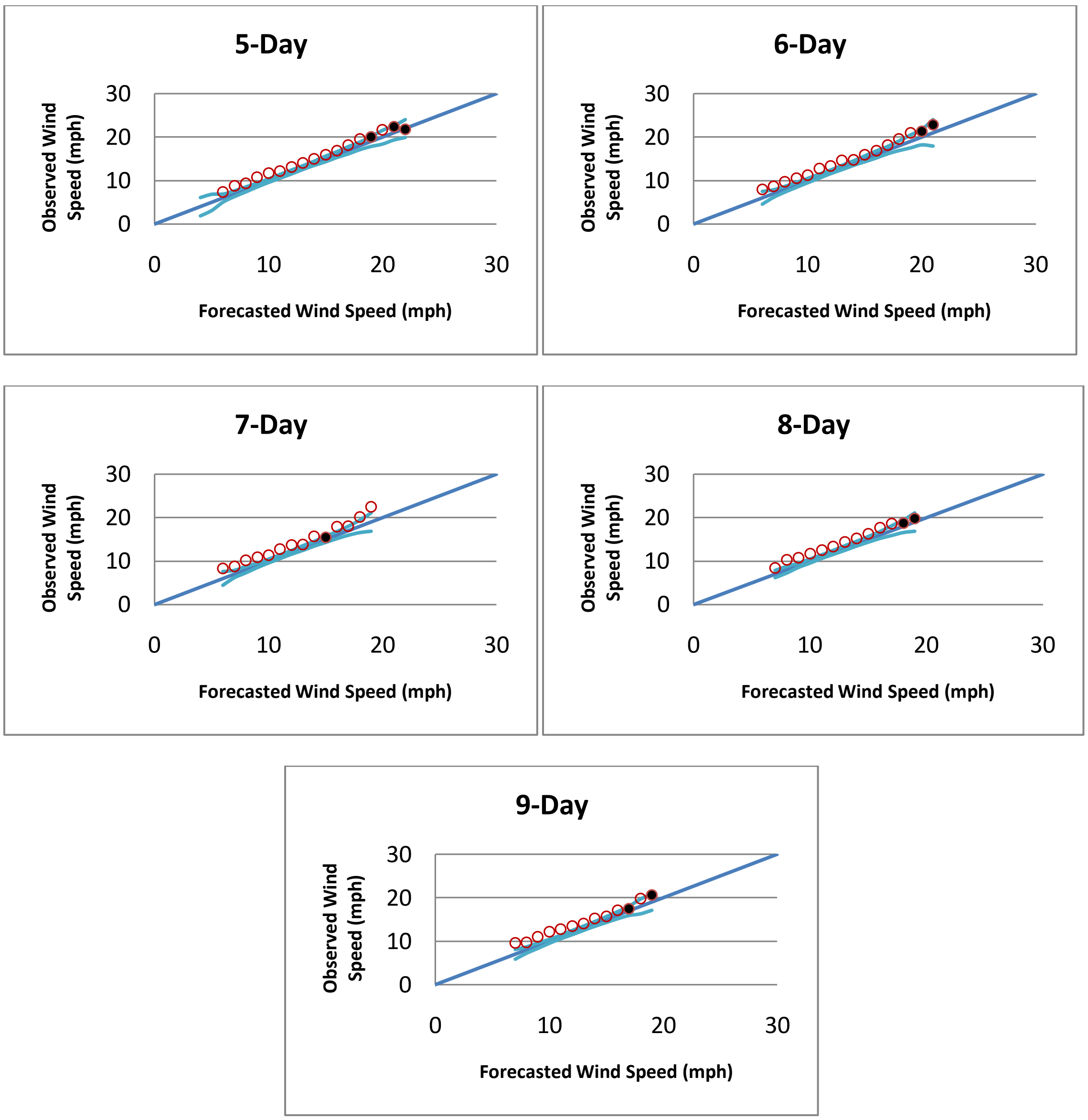


Figure A8b: Calibration Diagrams for TWC's 5-9 day lead times wind speed forecasts for Class 3 locations.

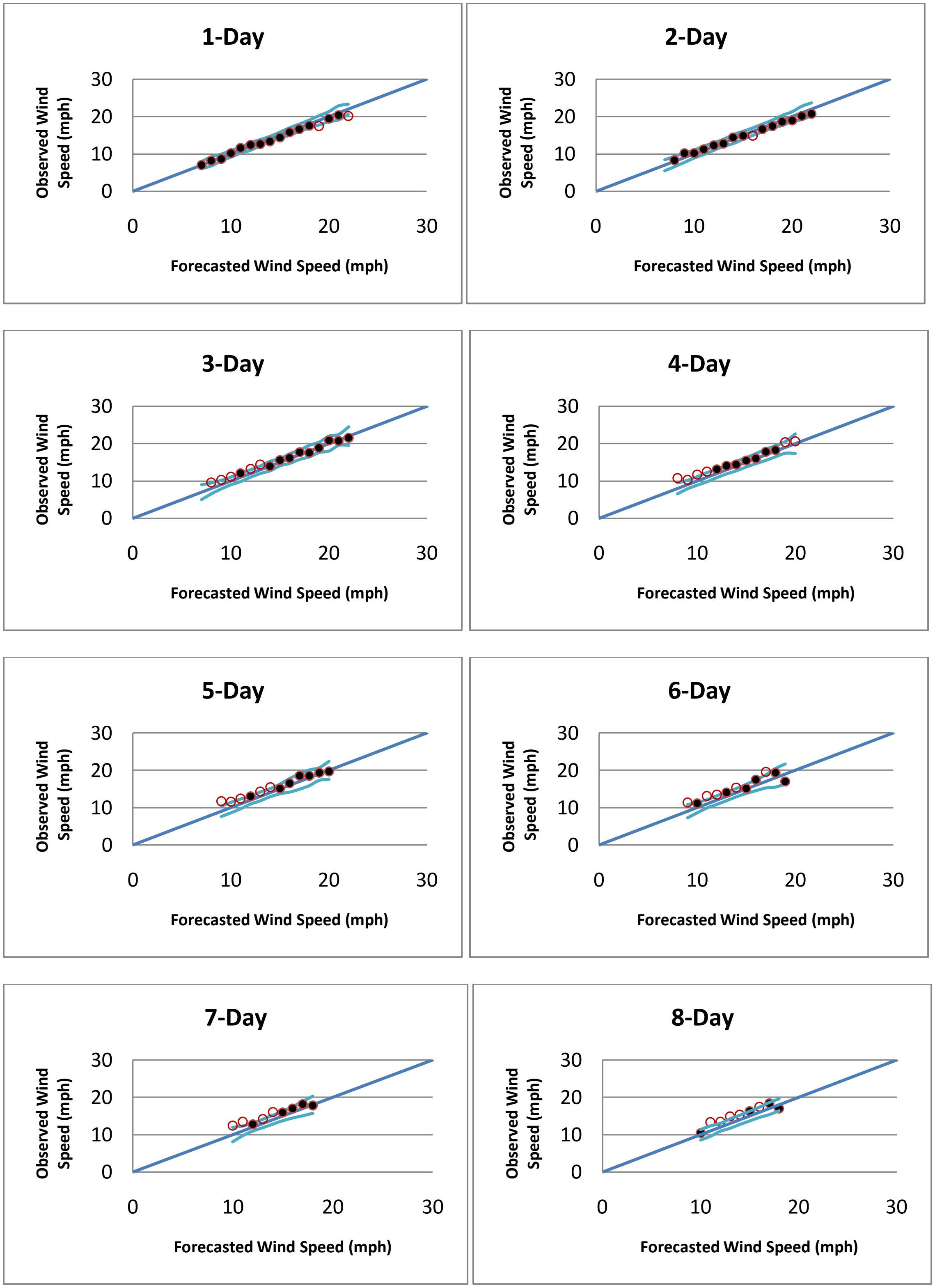


Figure A9a: Calibration Diagrams for TWC's 1-8 day lead times wind speed forecasts for Class 4 locations.

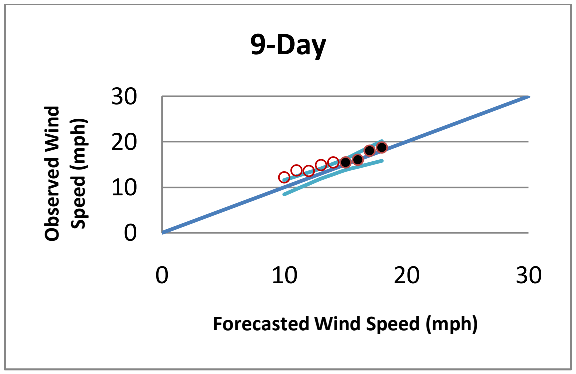


Figure A9b: Calibration Diagrams for TWC's 9 day lead time for Class 4 locations.

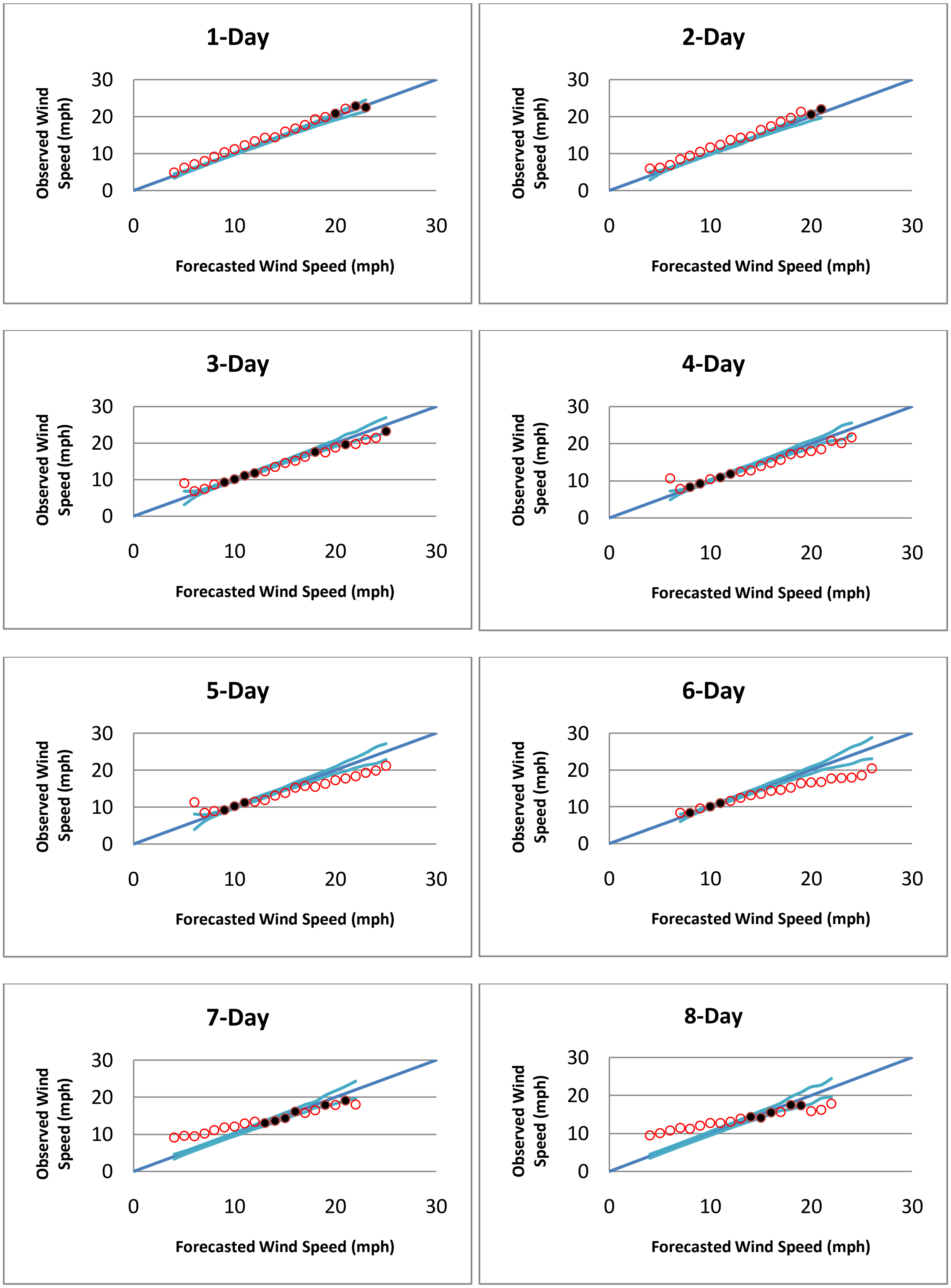


Figure A10a: Calibration Diagrams for AWX's 1-8 day lead times for Class 2 locations.

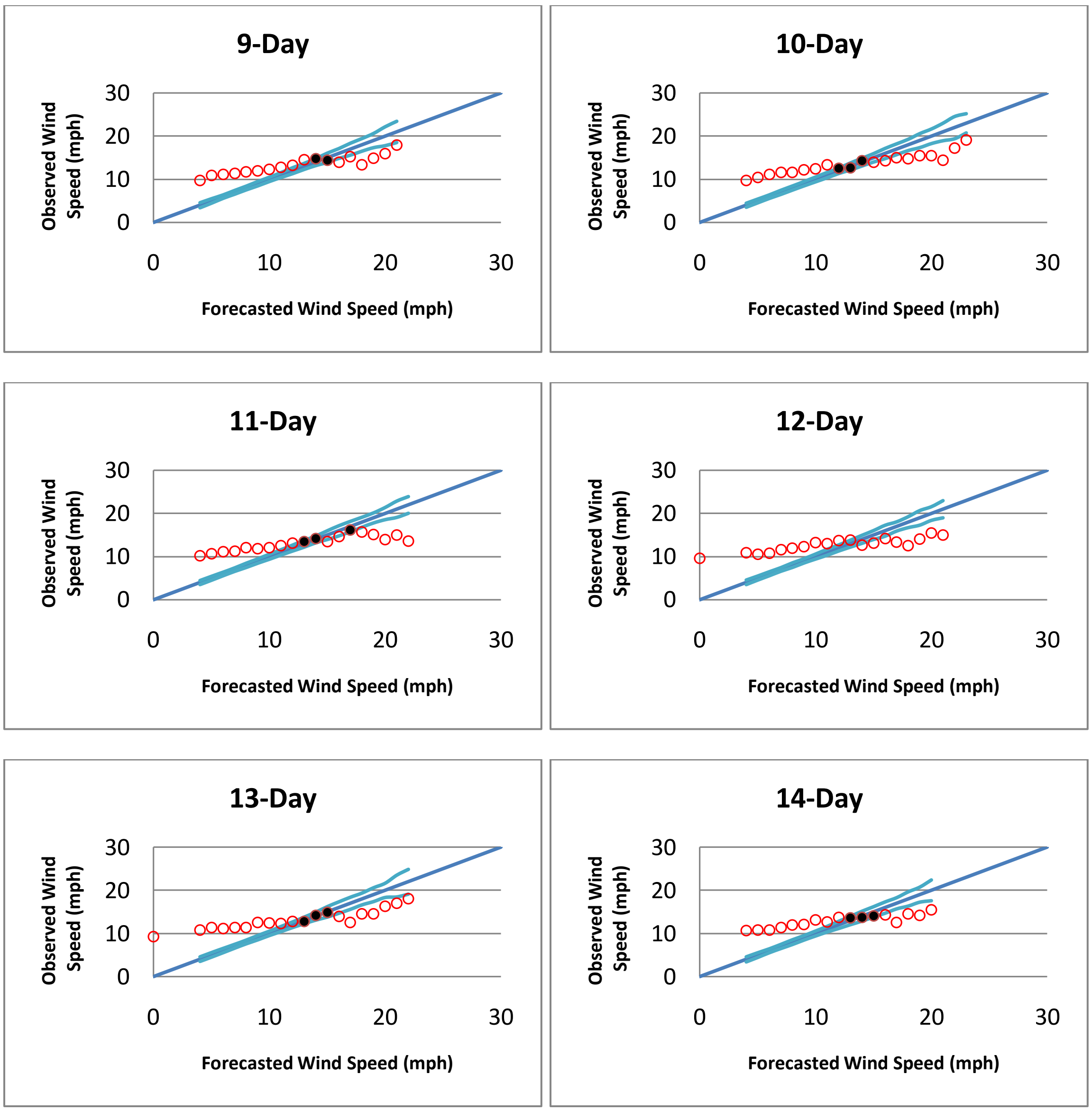


Figure A10b: Calibration Diagrams for AWX's 9-14 day lead times for Class 2 locations.

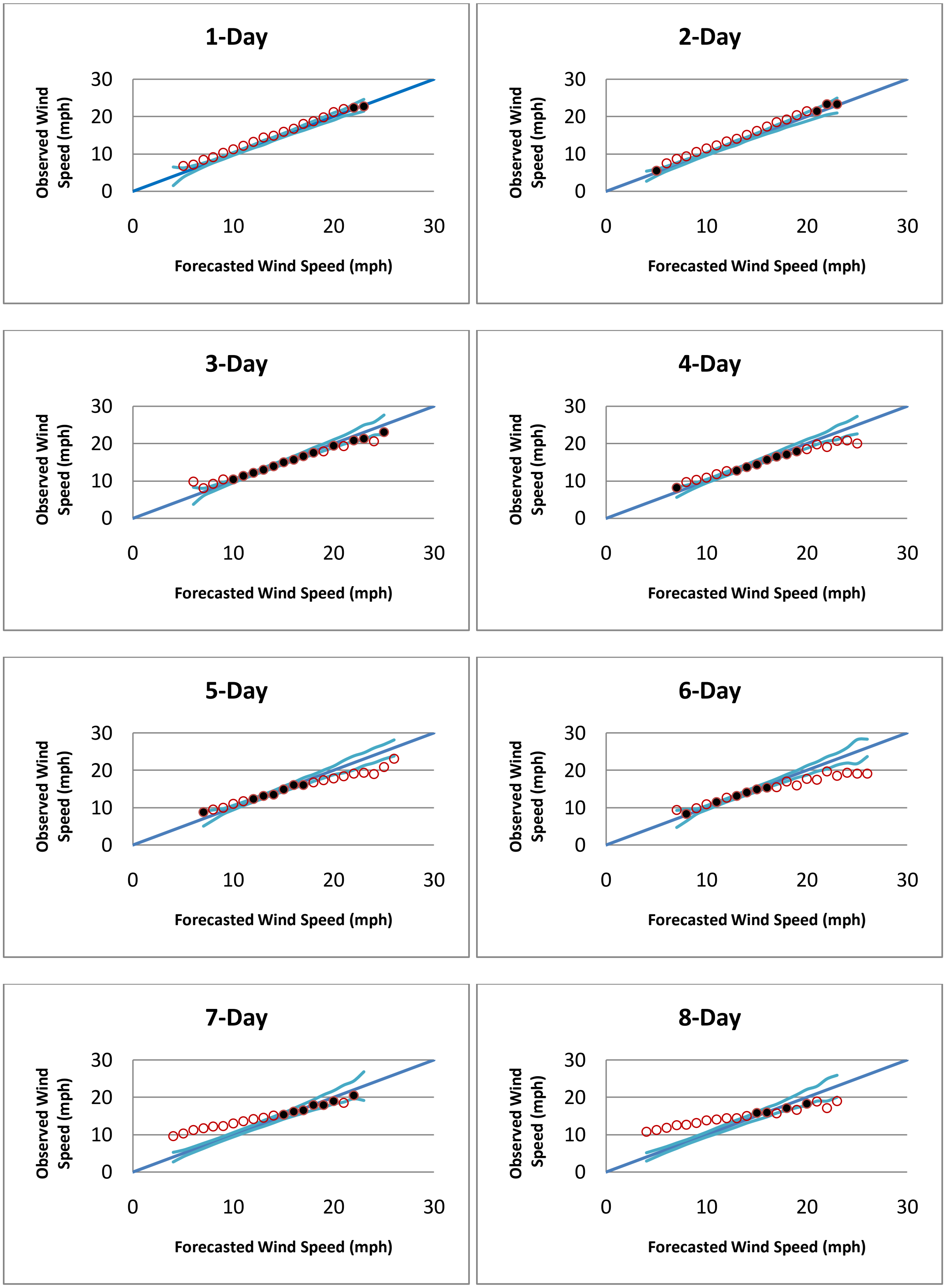


Figure A11a: Calibration Diagrams for AWX's 1-8 day lead times for Class 3 locations.

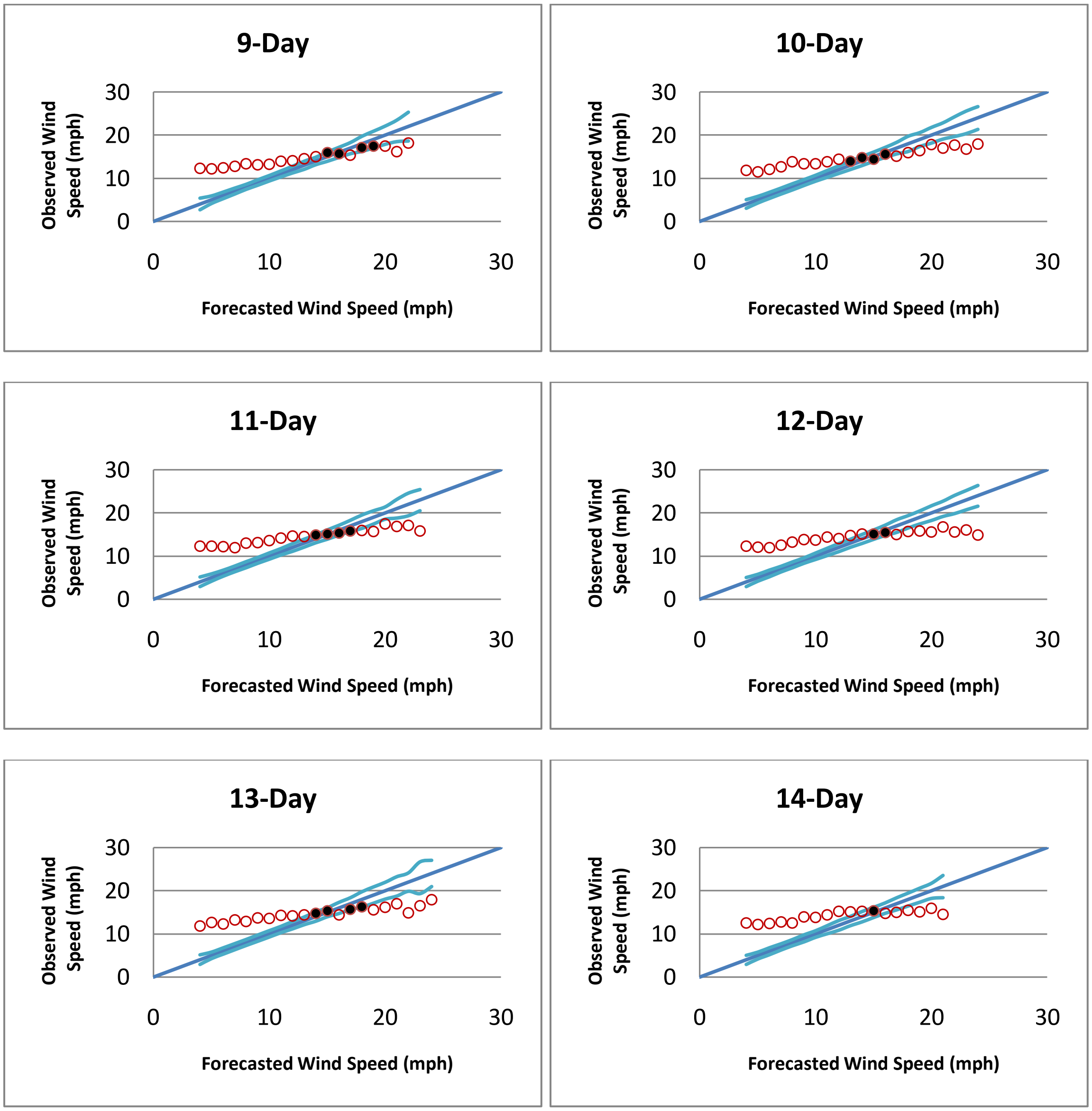


Figure A11b: Calibration Diagrams for AWX's 9-14 day lead times for Class 3 locations.

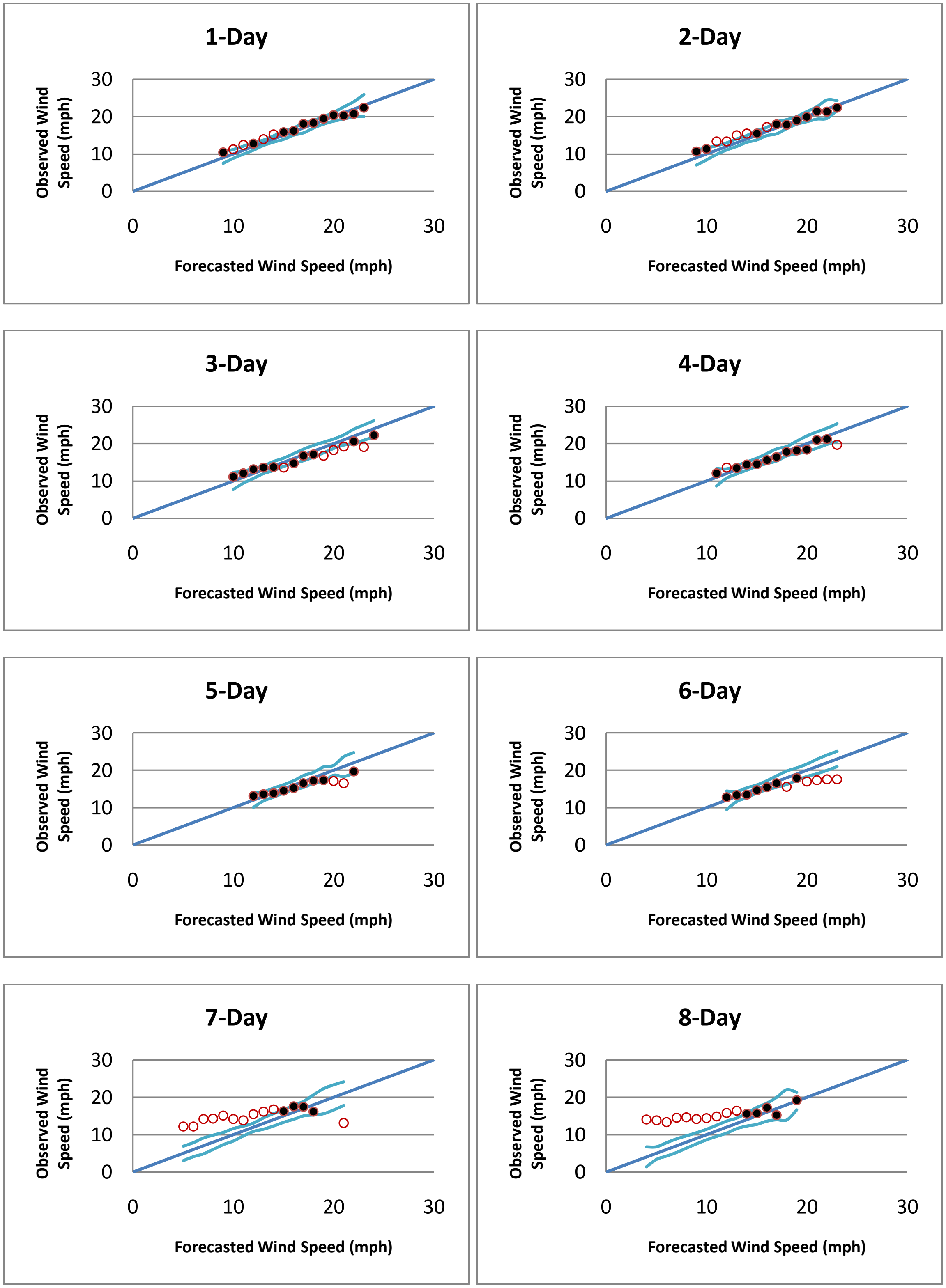


Figure A12a: Calibration Diagrams for AWX's 1-8 day lead times for Class 4 locations.

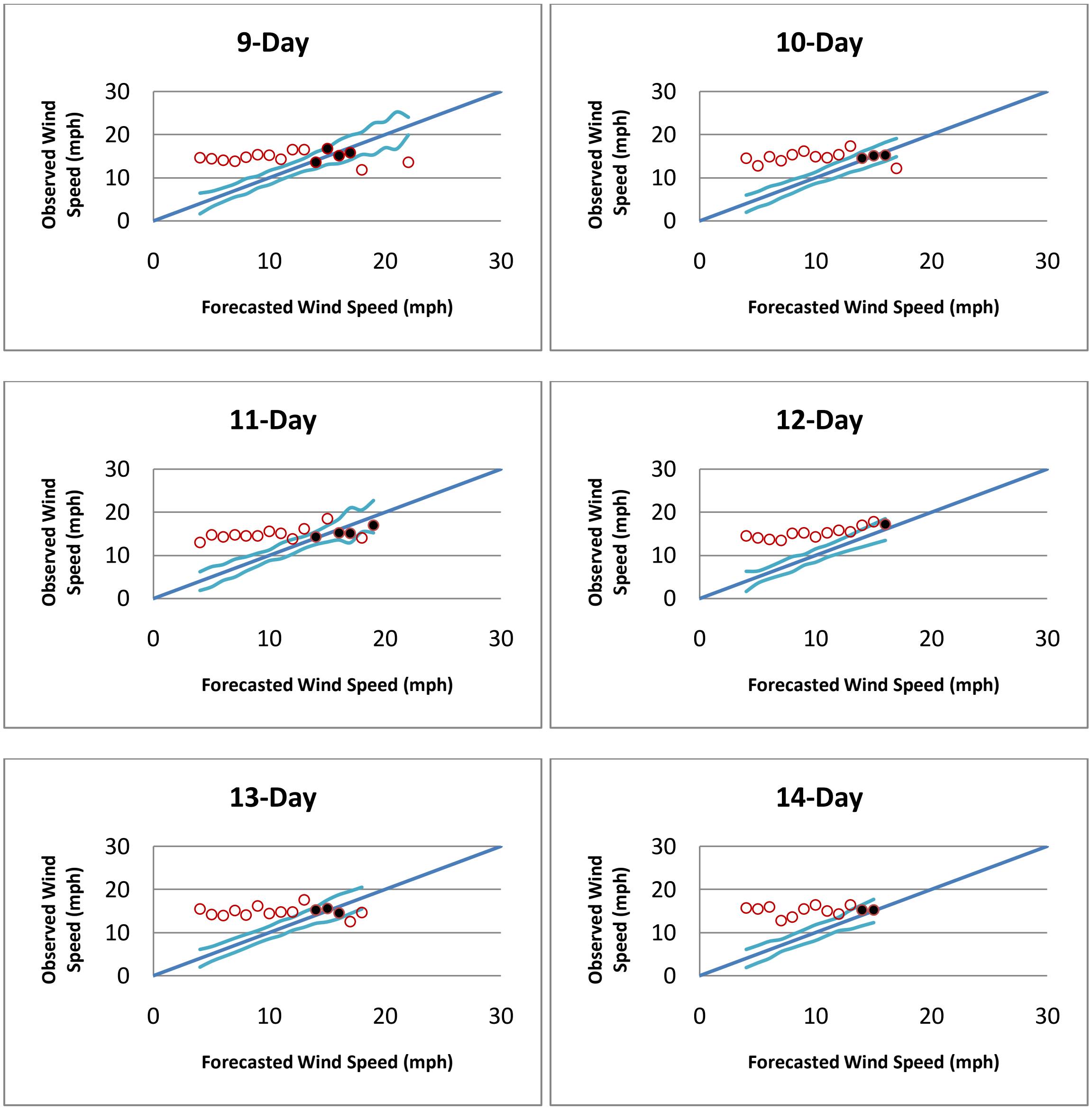


Figure A12b: Calibration Diagrams for AWX's 9-14 day lead times for Class 4 locations.

References

Accadia, C., S. Zecchetto, A. Lavagnini, A. Speranza, 2007: Comparison of 10-m Wind Forecasts from a Regional Area Model and QuikSCAT Scatterometer Wind Observations over the Mediterranean Sea. *Mon. Wea. Rev.,* **135**, 1945-1960.

Bickel, J.E. and Kim, S.D., 2008: Verification of The Weather Channel Probability of Precipitation Forecasts. *Monthly Weather Review,* **136**, 4867-4881.

Elliott, D.L., C.G. Holladay, W.R. Barchet, H.P. Foote, W.F. Sandusky. *Wind Energy Resource Atlas of the United States*. Retrieved May 15, 2011, from http://rredc.nrel.gove/pubs/atlas/

Feddersen, H., K. Sattler, 2005: Verification of wind forecasts for a set of experimental DMI-HIRLAM ensemble experiments. Danish Meteorological Institute Scientific Report.

Finley, J.P., 1884. Tornado Predictions. *Amer. Met. J.,* **1,** 85-88.

Jolliffe, I.T. and D.B. Stephenson, Eds., 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science.* John Wiley and Sons, 240 pp.

Milligan, M.R., A.H. Miller, F. Chapman, 1995: Estimating the Economic Value of Wind Forecasting to Utilities. *NREL: National Renewable Energy Laboratory.* Retrieved June 5, 2011, from http://www.nrel.gov/wind/systemsintegration/pdfs/1995/ milligan_wind_forecasting.pdf

Murphy, A.H. and Winkler, R.L., 1984. A general framework for forecast verification. *Mon. Weather Rev.,* **115,** 1330-1338.

Murphy, A.H. and Epstein, E.S., 1989. A note on probability forecasts and 'hedging'. *J. Appl. Met.,* **117**, 572-581.

White, B.G., J. Paegle, W.J. Steenburgh, J.D. Horel, R.T. Swanson, L.K. Cook, D.J. Onton, J.G. Miles, 1998: Short-Term Forecast Validation of Six Models. *Wea. Forecasting*, **14**, 84-108.